

University of Groningen

The threefold evaluation of theories

Kuipers, Theo A.F.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2001

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kuipers, T. A. F. (2001). *The threefold evaluation of theories: a synopsis of from instrumentalism to constructive realism. On some relations between confirmation, empirical progress, and truth approximation* (2000). s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

THE THREEFOLD EVALUATION OF THEORIES

A Synopsis of

"From Instrumentalism to Constructive Realism. On Some Relations between Confirmation, Empirical Progress, and Truth Approximation" (2000)

Abstract

Surprisingly enough, modified versions of the confirmation theory of Carnap and Hempel and the truth approximation theory of Popper turn out to be smoothly synthesizable. The glue between confirmation and truth approximation appears to be the instrumentalist methodology, rather than the falsificationist one.

By evaluating theories separately and comparatively in terms of their successes and problems (hence even if they are already falsified), the instrumentalist methodology provides – both in theory and in practice – the straight route for short-term empirical progress in science in the spirit of Laudan. However, it is argued that such progress is also functional for all kinds of truth approximation: observational, referential, and theoretical. This sheds new light on the long-term dynamic of science and hence on the relation between the main epistemological positions, viz., instrumentalism (Toulmin, Laudan), constructive empiricism (van Fraassen), referential realism (Hacking and Cartwright), and theory realism of a non-essentialist nature (Popper), here called constructive realism.

In *From Instrumentalism to Constructive Realism* (2000) the above story is presented in great detail. The present synopsis highlights the main ways of theory evaluation presented in that book, viz. evaluation in terms of confirmation (or falsification), empirical progress and truth approximation.

Introduction

Over the years I have been working on two *prima facie* rather different, if not opposing, research programs, notably Carnap's confirmation theory and Popper's truth approximation theory. However, I have always felt that they must be compatible, even smoothly synthesizable, for all empirical scientists use confirmation intuitions, and many of them have truth approximation ideas. Gradually it occurred to me that the glue between confirmation and truth approximation was the instrumentalist or evaluation methodology, rather than the falsificationist one. By separate and comparative evaluation of theories in terms of their successes and problems – hence, even if already falsified – the evaluation methodology provides in theory and practice the straight route for short-term empirical progress in science in the spirit of Laudan. Further analysis showed that this also sheds new light on the long-term dynamic of science and hence on the relation between the main epistemological positions, viz., instrumentalism (Toulmin, Laudan), constructive empiricism (van Fraassen), referential realism (Hacking), and theory realism of a non-essentialist nature, here called constructive realism (Popper). Indeed, thanks to the evaluation methodology, there are good, if not strong reasons for all three epistemological transitions "from instrumentalism to constructive realism".

In this way a clear picture of scientific development arises, with a short-term and a long-term dynamic. In the former there is a severely restricted role for confirmation and falsification, the dominant role being played by (the aim of) empirical progress, and there are serious prospects for observational, referential and theoretical truth approximation. Hence, in regard to this short-term dynamic, the scientist's intuition that the debate among philosophers about instrumentalism and realism has almost no practical consequences can be explained and justified. The long-term dynamic is enabled by (observational, referential and theoretical) inductive jumps, after 'sufficient confirmation', providing the means to enlarge the observational vocabulary in order to investigate new domains of reality. In this respect, a

consistent instrumentalist epistemological attitude seems difficult to defend, whereas constructive realism seems most plausible.

In *From Instrumentalism to Constructive Realism* (ICR, 2000) the above story is presented in great detail. The present synopsis highlights the main ways of theory evaluation presented in ICR, viz. evaluation in terms of confirmation (or falsification), empirical progress and truth approximation. It essentially follows the division of ICR in four parts and 13 chapters, here resulting in four parts and 13 sections. However, Part IV (Chh. 10-12), dealing with refined truth approximation, is only briefly sketched. This synopsis is necessarily selective and hence it may be useful to consult from time to time the complete table of contents, including section titles, which is reproduced in Appendix 1.¹ Appendix 2 presents the outline table of contents of the companion volume *Structures in Science* (SiS, 2001), to which occasional reference will be made. Appendix 3 gives a list of acronyms.

I would like to conclude this introduction by referring to Ilkka Niiniluoto's major contribution to the field, viz. his *Truthlikeness* of 1987. Despite our differences regarding the topic of truth approximation, notably his emphasis on a quantitative approach and my emphasis on a qualitative one, the reader will come to understand that I feel much sympathy with his slogan "Popper's voice but Carnap's hands" (Niiniluoto, 1987, p. xvi).

1. General Introduction: Epistemological Positions

The core of the ongoing instrumentalism-realism debate concerns the nature of theoretical terms and of proper theories using such terms, or rather the attitude one should have towards them. *Prima facie*, the most important epistemological positions in that debate are certainly instrumentalism, constructive empiricism, referential realism and theory realism. They can be characterized and ordered according to the ways in which they answer a number of leading questions, where every subsequent question presupposes the affirmative answer to the previous one. For completeness, I start with two preliminary questions that get a positive answer from the major positions, but a negative one in idealist and extremely relativist postmodern circles:

Question 0: Does a natural world that is independent of human beings exist?

No: ontological idealism; Yes: ontological realism.

Question 1: Can we claim to possess true claims to knowledge about the natural world?

No: epistemological relativism; Yes: epistemological realism.

Question 2: Can we claim to possess true claims to knowledge about the natural world beyond what is observable?

No: empiricism: instrumentalism or constructive empiricism; Yes: scientific realism.

Question 3: Can we claim to possess true claims to knowledge about the natural world beyond (what is observable and) reference claims concerning theoretical terms?

No: entity or, more generally, referential realism; Yes: theory realism.

Question 4: Does there exist a correct or ideal conceptualization of the natural world?

¹ ICR is based on many publications, starting from 1978. The Foreword of ICR (p. x) mentions those 10 papers that have partially been used in writing ICR. This synopsis represents the main lines of ICR from its dominant point of view, viz. theory evaluation. It is supposed to be my last survey of ICR. A number of special topic-oriented surveys have been written before. Their titles indicate their special emphasis: Pragmatic aspects of truth approximation (1998), Abduction aiming at empirical progress or even truth approximation, leading to a challenge for computational modeling (1999), Progress in nomological, explicative and design research (Ch. 9 of SiS, 2001), Beauty, a road to the truth (2002), Empirical and conceptual idealization and concretization: the case of truth approximation (forthcoming a), Inference to the best theory: kinds of abduction and induction (forthcoming b).

No: constructive realism; Yes: essentialist realism.

Note first that 'empiricism' has two variants. They split on the subquestion whether reference of theoretical terms and truth values of theoretical statements even have to be formally denounced, notably as category mistakes by instrumentalists, or not, as constructive empiricism concedes. The splitting of 'theory realism' at the end of this question-and-answer game into 'constructive realism' and 'essentialist realism' suggests that we now have five main positions: instrumentalism, constructive empiricism, and referential, constructive and essentialist realism. The following scheme, starting with Question 2, presents their relation in brief.

Q2:	true claims about the natural world beyond the observable?	⇒ no	empiricism - instrumentalism - constructive empiricism
	yes ↓ scientific realism		
Q3:	beyond reference?	⇒ no	referential realism ⇒ entity realism
	yes ↓ theory realism		
Q4:	ideal conceptualization?	⇒ no	constructive realism
	yes ↓ essentialist realism		

The main epistemological positions

Important refinements are obtained when the Questions 2 - 4 are considered from four perspectives on theories. On the one hand, theories supposedly deal only with 'the actual world' or primarily with 'the nomic world', that is, with what is possible in the natural world. On the other hand, one may only be interested in whether theories are true or false, or primarily in whether they approach 'the truth', regarding the world of interest. It should be stressed that 'the truth' is always to be understood in a domain-and-vocabulary relative way. Hence, no language-independent metaphysical or essentialist notion of 'the truth' is assumed. The four perspectives imply that all (non-relativistic) epistemological positions have an 'actual world version' and a 'nominal world version' and that they may be restricted to 'true-or-false' claims, or emphasize 'truth approximation claims'. In both cases it is plausible to distinguish between observational, referential, and theoretical claims and corresponding inductions, that is, the acceptance of such claims as true. Instrumentalists, in parallel, speak of theories as 'reliable-or-unreliable' derivation instruments or as 'approaching the best derivation instrument'. All four perspectives occur in particular in their realist versions, but they also make sense in adapted form in most of the other epistemological positions.

ICR is primarily a study of confirmation, empirical progress and truth approximation, and their relations. With the emphasis on their nomic interpretation the five main positions are further characterized and compared in the light of the results of this study, leading to the following conclusions. There are good reasons for the instrumentalist to become a constructive empiricist; in turn, in order to give deeper explanations of success differences, the constructive empiricist is forced to become a referential realist; in turn, there are good reasons for the referential realist to become a theory realist. The theory realist has good reasons to indulge in constructive realism, since there is no reason to assume that there are essences in the world, the existence of which is a prerequisite for ideal conceptualizations. As a result, the way leads to constructive realism and amounts to a pragmatic argument for this position, where the good reasons mainly deal with the short-term and the long-term dynamics

generated by the nature of, and the relations between, confirmation, empirical progress and truth approximation.

The suggested hierarchy of the heuristics corresponding to the epistemological positions is, of course, not to be taken in any dogmatic sense. That is, when one is unable to successfully use the constructive realist heuristic, one should not stick to it, but try weaker heuristics: first the referential realist, then the constructive empiricist, and finally the instrumentalist heuristic. For, as with other kinds of heuristics, although not everything goes all the time, *pace* (the suggestion of) Feyerabend's slogan "anything goes", everything goes sometimes. Moreover, after using a weaker heuristic, a stronger heuristic may become applicable at a later stage: "reculer pour mieux sauter".

Besides epistemological conclusions, there are some general methodological lessons to be drawn. The main one is that there are good reasons for all positions not to use the falsificationist but the instrumentalist or 'evaluation(ist)' methodology. That is, empirical (and to some extent perhaps non-empirical) successes and failures should exclusively guide the selection of theories, even if the better theory has already been falsified. This common methodology, directed at the separate and comparative evaluation of theories, is presented in Sections 5 and 6 below.

I CONFIRMATION

In this part a sketch of the main ideas behind confirmation and falsification of a hypothesis by the so-called HD(hypothetico-deductive) method is followed by a description of the 'landscape of qualitative and quantitative confirmation', as I like to call it. Confirmation of a hypothesis, however, has the connotation that the hypothesis has not yet been falsified. Whatever the truth claim associated with a hypothesis, as soon as it has been falsified, the plausibility (or probability) that it is true becomes and remains zero. In the next part I elaborate how theories can nevertheless be evaluated after falsification.

2. Confirmation by the HD Method

I start this section with a brief exposition of HD testing, that is, the HD method of testing hypotheses, and indicate the related qualitative explication of confirmation and, in the next section, its quantitative extensions. ICR Ch. 2, moreover, deals in detail with the paradoxes of the ravens and emeralds and some other confirmation problems and solutions. Moreover, it deals briefly with induction, that is, the acceptance of hypotheses.

HD testing

HD testing attempts to give an answer to one of the questions that one may be interested in, *the truth question*, which may be qualified according to the relevant epistemological position. The HD method prescribes the derivation of test implications and testing them. In each particular case, this may either lead to confirmation or to falsification. Whereas the 'language of falsification' is relatively clear, the 'language of confirmation' is a matter of great dispute.

According to the leading expositions of the hypothetico-deductive (HD) method by Hempel (1966), Popper (1934/1959) and De Groot (1961/1969), the aim of the HD method is to determine whether a hypothesis is true or false, that is, it is a method of testing. On closer inspection, this formulation of the aim of the HD method is not only laden with the epistemological assumption of theory realism, according to which it generally makes sense to aim at true hypotheses, but it also mentions only one of the realist aims, i.e., answering the 'truth question'. Applying the HD method to this end will be called HD testing as distinct from HD evaluation, which has other primary aims.

For the moment I will confine myself the attention to the HD method as a method of testing hypotheses. Though the realist has a clear aim in undertaking HD testing, this does not mean that HD testing is only useful from that epistemological point of view. Let me briefly

review the other main epistemological positions as far as the truth question is concerned and recall that claims may pertain to the actual world or to the nomic world (of nomic possibilities). Hypotheses may or may not use so-called 'theoretical terms' in addition to so-called 'observation terms'. What is observational is not taken in some absolute, theory-free sense, but depends greatly on the level of theoretical sophistication. Theoretical terms intended to refer to something in the actual or nomic world may or may not in fact successfully refer to something. For the (*constructive*) *empiricist* the aim of HD testing is to investigate whether the hypothesis is observationally true, i.e., has only true observational consequences, or is observationally or empirically adequate, to use Van Fraassen's favorite expression. For the *instrumentalist* the aim of HD testing is still more liberal (and essentially part of the aim of HD evaluation): for which intended applications is the hypothesis observationally true? The *referential realist*, on the other hand, adds to the aim of the empiricist to investigate whether the hypothesis is referentially true, i.e., whether its referential claims are correct. In contrast to the *theory realist*, he is not interested in the question whether the theoretical claims, i.e., the claims using theoretical terms, are true as well.

Methodologies are ways of answering epistemological questions. It turns out that the method of HD testing, the test methodology, is functional for answering the truth question of all four epistemological positions. For this reason, I shall present the test methodology in fairly neutral terms, viz., plausibility, confirmation and falsification.

The expression 'the plausibility of a hypothesis' abbreviates the informal qualification 'the plausibility, in the light of the background beliefs and the evidence, that the hypothesis is true'. Here 'true' may be specified in one of the four main senses: 1) observationally, as far as particular intended applications are concerned, 2) observationally, as far as all intended applications are concerned, 3) and, moreover, referentially, 4) and even theoretically. Admittedly, despite these possible qualifications, the notion of 'plausibility' remains necessarily vague, but that is what most scientists would be willing to subscribe to. When talking about 'the plausibility of certain evidence', I mean, of course, 'the prior plausibility of the (observational!) hypothesis that the test will result in the reported outcome'. Hence, here 'observationally true' and 'true' coincide by definition for what can be considered evidential statements.

Regarding the clarity of notions of 'confirmation' and 'falsification' the situation is rather asymmetric. 'Falsification' of a hypothesis simply means that the evidence entails that the hypothesis is observationally false, and hence also false in the stronger senses. However, what 'confirmation' of a hypothesis precisely means is not so clear. The explication of the notion of 'confirmation' of a hypothesis by certain evidence is here primarily approached from the success perspective on confirmation. This perspective equates confirmation with an increase in the plausibility of the evidence on the basis of the hypothesis, and implies that the evidence increases the plausibility of the hypothesis. However, by a liberalization suggested in debate with Maher (this volume), I now add "or an increase of the plausibility of the hypothesis on the basis of the evidence, if the latter has zero plausibility".

A test of a hypothesis may be experimental or natural. That is, a test may be an experiment, an active intervention in nature or culture, but it may also concern the passive registration of what is or was the case, or what happens or has happened. In the latter case of a so-called natural test, the registration may be a more or less complicated intervention, but is nevertheless supposed to have no serious effect on the course of events of interest.

According to the HD method a hypothesis H is tested by deriving test implications from it, and checking, if possible, whether they are true or false. Each test implication has to be formulated in terms that are considered to be observation terms. A test implication may or may not be general in nature. Usually there is background knowledge B , which is assumed to be true. Moreover, a test implication is usually of a conditional nature, if C then F ($C \rightarrow F$). Here C denotes one or more 'initial conditions' and F denotes a potential fact (event or state of affairs) predicted by H and C . If C and F are of an individual nature, F is called an individual test implication, and $C \rightarrow F$ a conditional test implication. When C is artificially realized, it is an experimental test, otherwise it is a natural test.

The basic logic of HD testing can be represented by some (valid) applications of Modus (Ponendo) Ponens (MP), where ' \models ' indicates 'logical entailment' and where ' I ' denotes a test implication:

$$\begin{array}{l} B, H \models I \\ \hline B, H \\ I \end{array} \qquad \begin{array}{l} B, H \models (C \rightarrow F) \\ \hline B, H, C \\ F \end{array}$$

It should be stressed that $B, H \models I$ and $B, H \models (C \rightarrow F)$ are supposed to be deductive claims, i.e., claims of a logico-mathematical nature.

The remaining logic of hypothesis testing concerns the application of Modus (Tollendo) Tollens (MT). Neglecting complications that may arise, such as that B 's or C 's truth may be disputed, if the test implication is false, the hypothesis must be false, and therefore has been falsified, for the following arguments are deductively valid (' \neg ' indicates 'negation'):

$$\begin{array}{l} B, H \models I \\ \hline B, \neg I \\ \neg H \end{array} \qquad \begin{array}{l} B, H \models (C \rightarrow F) \\ \hline B, C, \neg F \\ \neg H \end{array}$$

When the test implication turns out to be true, the hypothesis has of course not been (conclusively) verified, for the following arguments are invalid, indicated by '-/-':

$$\begin{array}{l} B, H \models I \\ B, I \\ -/-/- \\ H \end{array} \qquad \begin{array}{l} B, H \models (C \rightarrow F) \\ B, C, F \\ -/-/-/- \\ H \end{array}$$

Since the evidence (I or $C \& F$) is compatible with H , we may at least say that H may still be true. However, we can say more than that. Usually it is said that H has been confirmed. It is important to note that such confirmation by the HD method means more than mere compatibility; it is confirmation in the strong sense that H has obtained a *success* of a (conditional) deductive nature. By entailing the evidence, H makes the evidence as plausible as possible. This I call *the success perspective* on ((conditional) deductive) *confirmation*.

Falsification and confirmation have many complications, e.g., due to auxiliary hypotheses. I will deal with several complications, related to general and individual test implications, at the end of Section 5. As already indicated, however, there is a great difference between falsification and confirmation. Whereas the 'logical grammar' of falsification is not very problematic, the grammar of confirmation, i.e., the explication of the concept of confirmation, has been a subject of much dispute.

Deductive confirmation

The grammar of confirmation to be presented in this and the following two sections (based on SiS, Subsection 7.1.2, and introducing some new elements relative to ICR) is in many respects a systematic exposition of well-known ideas about deductive, structural, and inductive confirmation. However, these ideas are presented in a non-standard way and refine and revise several standard solutions of problems associated with these ideas.

Here I will only give a sketch of the main lines of the three ICR chapters on confirmation. It is important to note that, although the role of falsification and confirmation will be relativized in many respects in part II, it will also become clear that they remain very important for particular types of hypotheses. Notably, they remain relevant for general observational (conditional) hypotheses, and for several kinds of (testable) comparative hypotheses, e.g., hypotheses claiming that one theory is more successful or (observationally, referentially or theoretically) even more truthlike than another.

This section deals with qualitative (deductive) confirmation that results from applying the HD method, while the next one deals with quantitative, more specifically probabilistic confirmation, including a suitable degree of confirmation. The third section introduces the crucial distinction between structural and inductive confirmation and gives a brief survey of the main systems of inductive confirmation in the Carnap-Hintikka tradition of so-called inductive logic, with a suitable degree of *inductive* confirmation.

The main non-standard aspect is the approach of confirmation from the 'success perspective', according to which confirmation is primarily equated with evidential success, more specifically with an increase of the plausibility of the *evidence* on the basis of the hypothesis. Hence, in contrast to standard expositions, confirmation is not directly (at least not in general, see below) equated with an increase of the plausibility of the *hypothesis* by the evidence. This is merely an additional aspect of confirmation under appropriate conditions and epistemological assumptions.

Contrary to many critics, I believe that the notion of *deductive (d-)confirmation* makes perfectly good sense, provided the classificatory definition is supplemented with some comparative principles. More specifically, "(contingent) evidence E d-confirms (consistent) hypothesis H , assuming B " is defined by the clause: $B \& H$ (logically) entails E , and further obeys:

Comparative principles:

P1: if $B \& H$ entails E and E entails E^* (and not vice versa) then E d-confirms H , assuming B , more than E^* .

P2: if $B \& H$ and $B \& H^*$ both entail E then E d-confirms H and H^* , assuming B , equally.

To be sure, this definition-with-comparative-supplement only makes sense as a *partial* explication of the intuitive notion of confirmation; it leaves room for non-deductive, in particular probabilistic extensions, as we shall see below. However, let us first look more closely at the comparative principles, suppressing the phrase "assuming B ". They are very reasonable in the light of the fact that the deductive definition can be conceived as a (deductive) *success* definition of confirmation: if H entails E , E clearly is a success of H , if not a predictive success, then at least a kind of explanatory success. From this perspective, P1 says that a stronger (deductive) success confirms a hypothesis more than a weaker one, and P2 says that two hypotheses should equally be praised for the same success. In particular P2 runs counter to standard conceptions. However, in Chapter 2 of ICR I deal extensively with the possible objections and show, moreover, that the present analysis can handle the confirmation paradoxes discussed by Hempel and Goodman.

I would like to conclude this section with the 'Confirmation Matrix', i.e., a survey of the four logical relations, with epistemologically plausible names, between a hypothesis and evidence, assuming background knowledge. Recall that 'd-' is short for 'deductive(ly)', '¬' indicates 'negation' and '≡' indicates 'logical entailment'. E is assumed to be true.

Conclusion Premises	E (true)	$\neg E$ (false)
B, H	E d-confirms H , assuming B $B \& H \equiv E$	E falsifies H , assuming B $B \& H \equiv \neg E$ ($\Leftrightarrow B \& E \equiv \neg H$)
$B, \neg H$	E d-disconfirms H , assuming B $B \& \neg H \equiv E$	E verifies H , assuming B $B \& \neg H \equiv \neg E$ ($\Leftrightarrow B \& E \equiv H$)

We get conditional notions by also assuming, besides background knowledge B , one or more initial conditions C , which play(s) logically the same role as B . Although C may not entail E , E may or may not be formulated such that it entails C .

3. Quantitative Confirmation, And Its Qualitative Consequences

There is a natural quantitative refinement of deductive confirmation, which will here be characterized. ICR Ch. 3 presents in detail several qualitative consequences, and also discusses the prospects for quantitative acceptance criteria for hypotheses. In the first appendix the theory of quantitative confirmation is compared with Popper's theory of corroboration, in the second its solution of the ravens paradoxes is compared with the standard Horwich' analysis.

Probabilistic confirmation presupposes, by definition, a probability function, indicated by p , that is, a real-valued function obeying the standard axioms of probability, which may nevertheless be of one kind or another (see Section 4). But first I shall briefly deal with the general question of a probabilistic criterion of confirmation and a degree of confirmation.

The *standard* (or *forward*) *criterion* for probabilistic confirmation is that the *posterior* probability $p(H/E)$ exceeds the (relative to the background knowledge) *prior* probability $p(H)$, that is, $p(H/E) > p(H)$. However, this criterion is rather inadequate for ' p -zero' hypotheses. For example, if $p(H)=0$ and E d-confirms H , this confirmation cannot be seen as an extreme case of probabilistic confirmation, since $p(H/E)=p(H)=0$. In other words, this criterion has the very undesirable consequence that p -zero hypotheses cannot be confirmed. However, for p -non-zero hypotheses and assuming $0 < p(E) < 1$, the standard criterion is equivalent to the *backward* or *success* criterion, according to which the so-called *likelihood* $p(E/H)$ exceeds the initial probability $p(E)$ of E : $p(E/H) > p(E)$.² Now it is easy to check that any probability function respects d-confirmation according to this criterion, since $p(E/H) = 1$ when H entails E , and hence exceeds $p(E)$, even if $p(H)=0$, in which case it is a matter of a plausible definition. More generally, the success criterion can apply in all p -zero cases in which $p(E/H)$ can nevertheless be meaningfully interpreted.

To be sure, as Maher (this volume) stresses, the success criterion does not work properly for ' p -zero evidence', e.g. in the case of verification of a real-valued interval hypothesis by a specific value within that interval. However, although this is less problematic than the case of p -zero hypotheses dealt with above (see my reply to Maher), it seems reasonable to accept the standard criterion for p -zero evidence. From now on 'confirmation' will mean forward or backward confirmation when $p(H) \neq 0 \neq p(E)$, backward confirmation when $p(H)=0$ and $p(E) \neq 0$ and forward confirmation when $p(E)=0$ and $p(H) \neq 0$; and it is left undefined when $p(H)=0=p(E)$.

In sum, we obtain the following survey of deductive confirmation, 'general confirmation', or simply 'confirmation', and related notions

E d-confirms H , assuming B	$B \& H \models E$	and hence $p(E/B \& H)=1$
E confirms H , assuming B	$p(E/B \& H) > p(E/B)$	or $p(H/B \& E) > p(H/B)$
E falsifies H , assuming B	$B \& H \models \neg E$	and hence $p(E/B \& H)=0$
or, equivalently,	$B \& E \models \neg H$	and hence $p(H/B \& E)=0$
E is neutral wrt H , assuming B	$p(E/B \& H) = p(E/B)$ and $p(H/B \& E)=p(H/B)$	
E d-disconfirms H , assuming B	$B \& \neg H \models E$	and hence $p(E/B \& \neg H)=1$
E disconfirms H , assuming B	$p(E/B \& H) < p(E/B)$	or $p(H/B \& E) < p(H/B)$
E verifies H , assuming B	$B \& \neg H \models \neg E$	and hence $p(\neg E/B \& \neg H)=1$
or, equivalently,	$B \& E \models H$	and hence $p(H/B \& E)=1$

² This follows directly from the general definition of conditional probability, viz., $p(A/B) =_{\text{def}} p(A \& B)/p(B)$, assuming that $p(B) \neq 0$. Note that this definition creates in general a tension between cases where $p(B)=0$ while we would also like to say that $p(A/B)=1$ because of the fact that B entails A .

Finally, we can define 'non-deductive confirmation' as 'confirmation, but no d-confirmation', and 'proper confirmation' as 'confirmation, but no verification'.

I now turn to the definition of a degree of confirmation, suppressing "assuming B". I propose, instead of the standard difference measure $p(H/E) - p(H)$, the non-standard ratio measure $p(E/H)/p(E)$ as the degree (or rate) of (backward) confirmation (according to p), indicated by $c_p(H, E)$. This ratio has the following properties. For p -non-zero hypotheses it is equal to the standard ratio measure $p(H/E)/p(H)$, and hence is symmetric ($c_p(H, E) = c_p(E, H)$), for p -non-zero hypotheses, but it leaves room for confirmation (amounting to: $c_p(H, E) > 1$) of p -zero-hypotheses. For p -zero evidence we may turn to the standard ratio measure.

The definition satisfies the comparative principles of deductive (d-)confirmation P1 and P2. Note first that $c_p(H, E)$ is equal to $1/p(E)$ when H entails E , for $p(E/H)=1$ in that case. This immediately implies P2: if H and H^* both entail E then $c_p(H, E) = c_p(H^*, E)$. Moreover, if H entails E and E^* , and E entails E^* (and not vice versa) then $c_p(H, E) > c_p(H, E^*)$, as soon as we may assume that $p(E) < p(E^*)$. Note that $p(E) \leq p(E^*)$ already follows from the assumption that E entails E^* . The result is a slightly weakened version of P1.

As suggested, there are a number of other degrees of confirmation. Fitelson (1999) evaluates four of them, among them the logarithmic forward version of my backward ratio measure, in the light of seven arguments or conditions of adequacy as they occur in the literature. The ratio measure fails in five cases. Three of them are directly related to the 'pure' character of r , that is, its satisfaction of P2.³ P2 is defended extensively in Chapter 2 of ICR.

However, I also argue there, in Chapter 3, that as soon as one uses the probability calculus, it does not matter very much which 'confirmation language' one chooses, for that calculus provides the crucial means for updating the plausibility of a hypothesis in the light of evidence. Hence, the only important point, which then remains, is to always make clear which confirmation language one has chosen.

4. Inductive Confirmation and Inductive Logic

I now turn to a discussion of some possible kinds of probability functions and corresponding kinds of probabilistic confirmation. ICR Ch. 4, moreover, deals in detail with Carnap's and Hintikka's theories of inductive confirmation of singular predictions and universal generalizations. It further summarizes some of the main results, following Carnap, that have been obtained with respect to optimum inductive probability functions and inductive analogy by similarity and proximity.

Structural confirmation

I start with *structural confirmation*, which has an objective and a logical version. Consider first an example dealing with a fair die. Let E indicate the even (elementary) outcomes 2, 4, 6, and H the 'high' outcomes 4, 5, 6. Then (the evidence of) an even outcome confirms the hypothesis of a high outcome according to both criteria, since $p(E/H)=p(H/E) = 2/3 > 1/2 = p(H)=p(E)$. I call this the paradigm example of (non-deductive) structural confirmation.

This example illustrates what Salmon (1969) already pointed out in the context of discussing the possibilities of an inductive logic. A probability function may be such that E confirms H in the sense that H *partially* entails E . Here 'partial entailment' essentially amounts to the claim that the relative number of models in which E is true on the condition that H is true is larger than the relative number of models in which E is true without any condition. This general idea can be captured in a quantitative way by defining structural confirmation as (backward) confirmation based on a probability function assigning constant

³ The P2-related arguments concern the first and the second argument in Fitelson's Table 1, and the second in Table 2. Of the other two, the example of 'unintuitive' confirmation is rebutted in ICR (Chapter 3) with a similar case against the difference measure. The other one is related to the 'grue-paradox', for which Chapter 2 and 3 of ICR claim to present an illuminating analysis in agreement with P2.

(e.g. equal) probabilities to the elementary outcomes. Such a probability function may represent an *objective* probability process, such as a fair die. Note that in this paradigm example of structural confirmation, that is, an even outcome of a fair die confirms a high outcome, the corresponding degree of confirmation is $(2/3)/(1/2)=4/3$. The probability function may also concern the so-called *logical* probability or logical measure function (Kemeny, 1953), indicated by m . Kemeny's m -function assigns probabilities on the basis of ((the limit of) the ratio of) the number of structures making a proposition true, that is, its number of models (cf. the random-world or labeled method in Grove, Halpern and Koller, 1996). These logical probabilities may or may not correspond to the objective probabilities of an underlying process, as is the case with a fair die. Hence, for structural confirmation, we may restrict the attention to (generalizations of) Kemeny's m -function.

Structural confirmation is a straightforward generalization of d-confirmation. For suppose that H entails E . Then $m(E/H) = (\lim) \text{Mod}(E\&H) / \text{Mod}(H) = 1 > (\lim) \text{Mod}(E) / \text{Mod}(\text{Tautology}) = m(E)$, where e.g., ' $\text{Mod}(H)$ ' indicates the number of models of H . Moreover, as already indicated, it is a probabilistic explication of Salmon's (1969) idea of confirmation by 'partial entailment', according to which an even outcome typically is partially implied by a high outcome.

It is important to note that the m -function leads in many cases to ' m -zero' hypotheses (cf. Compton, 1988). For instance, every universal generalization "for all x Fx " gets zero m -value for an infinite universe. As we may conclude from the general exposition in Section 3, certain evidence may well structurally confirm such hypotheses by definition, according to the success criterion, but not according to the standard criterion. E.g., a black raven structurally (conditionally deductively) confirms "all ravens are black" according to the success criterion, even if the universe is supposed to be infinite. In this case the m -value of that hypothesis is zero, with the consequence that it is not confirmed according to the standard criterion. However, it is typical for the m -function that it lacks, even from the success perspective, the confirmation property which is characteristic of inductive probability functions.

Inductive confirmation

Inductive confirmation is (pace Popper and Miller, 1983) explicated in terms of confirmation based on an inductive probability function, i.e., a probability function p having the general feature of 'positive relevance', 'inductive confirmation' or, as I like to call it,

$$\text{instantial confirmation: } p(Fa/E\&Fb) > p(Fa/E)$$

where ' a ' and ' b ' represent distinct individuals, ' F ' an arbitrary monadic property and ' E ' any kind of contingent evidence compatible with Fa . Note that this definition is easy to generalize to n -tuples and n -ary properties, but I will restrict the attention to monadic ones. Since the m -function satisfies the condition $m(Fa/Fb\&E) = m(Fa/E)$, we get for any inductive probability function p :

$$p(Fa\&Fb/E) = p(Fa/E).p(Fb/E\&Fa) > m(Fa\&Fb/E)$$

Inductive (probability) functions can be obtained in two ways, which may also be combined:

- 'inductive priors', i.e., positive prior p -values $p(H)$ for m -zero hypotheses

and/or

- 'inductive likelihoods', i.e., likelihood functions $p(E/H)$ having the property of instantial confirmation

Note first that forward confirmation of m -zero hypotheses requires inductive priors, whereas backward confirmation of such hypotheses is always possible, assuming that $p(E/H)$ can be

interpreted. Below I will give a general definition of inductive confirmation in terms of degrees of confirmation.

In terms of the two origins of inductive probability functions we can characterize the four main theories of confirmation in philosophy of science as follows:

	Inductive priors	Inductive likelihoods
Popper	No	No
Carnap	No	Yes
Bayes ⁴	Yes	No
Hintikka	Yes	Yes

Popper rejected both kinds of inductive confirmation, for roughly three reasons: two problematic ones and a defensible one. The first problematic one (Popper, 1934/1959) is that he tried to argue, not convincingly (see e.g., Earman, 1992, Howson and Urbach, 1989, Kuipers, 1978), that $p(H)$ could not be positive. The second one is that any probability function has the property " $p(E \rightarrow H/E) < p(E \rightarrow H)$ " (Popper and Miller, 1983). Although the claimed property is undisputed, the argument that a proper inductive probability function should have the reverse property, since ' $E \rightarrow H$ ' is the 'inductive conjunct' in the equivalence " $H \Leftrightarrow (E \vee H) \& (E \rightarrow H)$ ", is not convincing. The indicated reverse property may well be conceived as an unlucky first attempt to explicate the core of (probabilistic) inductive intuitions, which should be replaced by the property of instancial confirmation. The defensible reason is that the latter property merely reflects a subjective attitude and, usually, not an objective feature of the underlying probability process, if there is such a process at all.

Carnap, following Laplace, favored inductive likelihoods, although he did not reject inductive priors. The so-called Bayesian approach in philosophy of science reflects inductive priors (but Bayesian statistics uses inductive likelihoods as well, see Festa, 1993). Finally, Hintikka introduced 'double inductive' probability functions, by combining the Carnapian and the Bayesian approach.

Degree of (inductive) confirmation

I now turn to the problem of defining a *degree of inductive confirmation* such that it entails a general definition of *inductive confirmation*. The present approach is not in the letter but in the spirit of Mura (1990) (see also e.g., Schlesinger, 1995) and Milne (1996) and Festa (1999). The idea is to specify a measure for the degree of inductive influence by comparing the relevant '*p*-expressions' with the corresponding (structural) '*m*-expressions' in an appropriate way. I proceed in two stages.

Stage 1. In the first stage we define, as announced, the degree of inductive influence in this degree of confirmation, or simply the degree of inductive (backward) confirmation (according to p), as the ratio:

$$r_p(H, E) = \frac{c_p(H, E)}{c_m(H, E)} = \frac{p(E/H)/p(E)}{m(E/H)/m(E)}$$

A direct consequence of this definition is that the degree of confirmation equals the product of the degree of structural confirmation and the degree of inductive confirmation.

Stage 2. In the second stage I generally define *inductive confirmation*, that is, E inductively confirms H , of course, by the condition: $r_p(H, E) > 1$. This definition leads to four interesting possibilities for confirmation according to p . Assume that $c_p(H, E) > 1$. The *first* possibility is *purely structural* confirmation, that is, $r_p(H, E) = 1$, in which case the confirmation has *no inductive features*. This trivially holds in general for structural confirmation, but it may occasionally apply to cases of confirmation according to some p

⁴ Here 'Bayes' refers to the Bayesian confirmation theory in the Howson-Urbach style, see below, not to Bayesian statistics.

different from m . The *second* possibility is that of *purely inductive* confirmation, that is, $c_m(H, E) = 1$, and hence $r_p(H, E) = c_p(H, E)$. This condition typically applies in the case of *instantial* confirmation, since, e.g., $m(Fa/Fb \& E)/m(Fa/E) = 1$. The *third* possibility is that of a combination of structural and inductive confirmation: $c_m(H, E)$ and $c_p(H, E)$ both exceed 1, but the second more than the first. This type of *combined* confirmation typically occurs when a Carnapian inductive probability function is assigned e.g., in the case of a die-like object of which it may not be assumed that it is fair. Starting from equal prior probabilities for the six sides such a function gradually approaches the observed relative frequencies. Now suppose that among the even outcomes a high outcome has been observed more than expected on the basis of equal probability. In this case, (only) knowing in addition that the next throw has resulted in an even outcome confirms the hypothesis that it is a high outcome in two ways: structurally (see above) and inductively.

Example. Let n be the total number of throws so far, let n_i indicate the number of throws that have resulted in outcome i ($1, \dots, 6$). Then the Carnapian probability that the next throw results in i is $(n_i + \lambda/6)/(n + \lambda)$, for some fixed finite positive value of the parameter λ . Hence, the probability that the next throw results in an even outcome is $(n_2 + n_4 + n_6 + \lambda/2)/(n + \lambda)$, and the probability that it is 'even-and-high' is $(n_4 + n_6 + \lambda/3)/(n + \lambda)$. The ratio of the latter to the former is the posterior probability of a high next outcome given that it is even and given the previous outcomes. It is now easy to check that in order to get a degree of confirmation larger than the structural degree, which is $4/3$, as we have noted before, this posterior probability should be larger than the corresponding logical probability, which is $2/3$. This is the case as soon as $2n_2 < n_4 + n_6$, that is, when the average occurrence of '4' and '6' exceeds that of '2'.

Let me finally turn to the *fourth* and perhaps most surprising possibility: confirmation combined with the 'opposite' of inductive confirmation, that is, $r_p(H, E) < 1$, to be called *counter-inductive* confirmation. Typical examples arise in the case of deductive confirmation. In this case $r_p(H, E)$ reduces to $m(E)/p(E)$, which may well be less than 1. A specific example is the following: let E be $Fa \& Fb$ and let p be inductive then E d-confirms "for all x Fx " in a counter-inductive way. On second thoughts, the possibility of, in particular, deductive counter-inductive confirmation should not be surprising. Inductive probability functions borrow, as it were, the possibility of inductive confirmation by reducing the available 'amount' of deductive confirmation.

Further research will have to determine whether deductive and inductive confirmation can ever go together in a meaningful way. For the moment the foregoing completes the treatment of HD testing of a theory in terms of confirmation and falsification. I now turn to HD evaluation, which leaves room for continued interest in theories after their falsification.

II EMPIRICAL PROGRESS

HD testing attempts to give an answer to one of the questions in which one may be interested, *the truth question*, which may be qualified according to the relevant epistemological position. However, *the (theory) realist*, for instance, is not only interested in the truth question, but also in some other questions. To begin with, there is the more refined question of which (individual or general) facts the hypothesis explains (its explanatory successes) and which facts are in conflict with the hypothesis (its failures); *the success question* for short. I show in this part that the HD method can also be used in such a way that it is functional in (partially) answering this question. This method is called HD evaluation, and uses HD testing of test implications. Since the realist ultimately aims to approach the strongest true hypothesis, if any, i.e., *the (theoretical-cum-observational) truth* about the subject matter, the plausible third aim of the HD method is to help answer the question of how far a hypothesis is from the truth,

the truth approximation question. Here the truth will be taken in a relatively modest sense, viz., relative to a given domain and conceptual frame. In Section 7 I make plausible the contention that HD evaluation is also functional in answering the truth approximation question.

The other epistemological positions are guided by two related, but more modest success and truth approximation questions, and I shall show later that the HD method is also functional in answering these related questions. The *constructive empiricist* may not only be interested in the question of whether the theory is empirically adequate or observationally true; i.e., whether the observational theory implied by the full theory is true. He may also be interested in the refined success question about what its true observational consequences and its observational failures are, and in the question of how far the implied observational theory is from the strongest true observational hypothesis, *the observational truth*. The *referential realist* may, in addition, be interested in the truth of the reference claims of the theory and how far it is from the strongest true reference claim, *the referential truth*. The *instrumentalist* phrases the first question of the empiricist more liberally: for what (sub-)domain is it observationally true? He retains the success question of the empiricist. Finally, he will reformulate the third question as follows: to what extent is it the best (and hence the most widely applicable) derivation instrument?

The method of HD evaluation will turn out, in this part, to be a direct way to answer the success question and, in the next part, an indirect way to answer the truth approximation question, in both cases for all four epistemological positions. This part will again primarily be presented in a relatively neutral terminology, with specific remarks relating to the various positions. The success question will be presented in terms of successes and counterexamples⁵: what are the potential successes and counterexamples of the theory?

In sum, two related ways of applying the HD method to theories can be distinguished. The first one is *HD testing*, which aims to answer the truth question. However, as soon as the theory is falsified, the realist with falsificationist leanings, i.e., advocating exclusively the method of HD testing, sees this as a disqualification of an explanatory success. The reason is that genuine explanation is supposed to presuppose the truth of the theory. Hence, from the realist-falsificationist point of view a falsified theory has to be abandoned and one has to look for a new one.

The second method to be distinguished, *HD evaluation*, keeps taking falsified theories seriously. It tries to answer the success question, the evaluation of a theory in terms of its successes and counterexamples (problems) (Laudan, 1977). For the (non-falsificationist) realist, successes remain explanatory successes and, when evaluating a theory, they are counted as such, even if the theory is known to be false.

It is important to note that the term '(HD) evaluation' refers to the evaluation in terms of successes and counterexamples, and not in terms of truth approximation, despite the fact that the method of HD evaluation will nevertheless turn out to be functional for truth approximation. Hence, the method of HD evaluation can be used meaningfully without any explicit interest in truth approximation and without even any substantial commitment to a particular epistemological position stronger than instrumentalism.

Chapters 5 and 6 are pivotal in ICR by providing the glue between confirmation and truth approximation, for which reason they are here relatively extensively summarized. Moreover, they are also included in SiS (as Chh. 7 and 8). In addition to what I will present here, ICR Ch. 5 deals more in detail with 'falsifying general hypotheses', which if accepted lead to general problems of theories. Moreover, it briefly deals with statistical test implications. Anticipating Part III, ICR Ch. 6 indicates already why the evaluation methodology can be functional for truth approximation. Moreover it explains and justifies in greater detail than the present synopsis the non-falsificationist practice of scientists, as opposed to the explicit falsificationist view of many of them. This is not only the case in terms of fruitful dogmatism, as discovered by Kuhn and Lakatos, but also in terms of truth

⁵ If the reader finds that the term 'counterexample' has a realist, or falsificationist, flavor, it may be replaced systematically by 'problem' or 'failure'.

approximation, for example, by the paradigmatic non-falsificationist method of idealization and concretization, as propagated by Nowak.

5. Separate Evaluation of Theories by the HD Method

In this section it is shown that a decomposition of the HD method applied to theories is possible which naturally leads to an explication of the method of separate HD evaluation, using HD testing, even in terms of three models. Among other things, it will turn out that HD evaluation is effective and efficient in answering the success question. In the next section I use the separate HD evaluation of theories for their comparative HD evaluation.

Evaluation report

The core of the HD method for the evaluation of theories amounts to deriving from the theory in question, say X , General Test Implication (GTI's) and subsequently (HD) testing them. For every GTI I it holds that testing leads sooner or later either to a counterexample of I , and hence a counterexample of X , or to the (revocable) acceptance of I : a *success* of X . A counterexample, of course, implies the falsification of I and X . A success minimally means a 'derivational success'; it depends on the circumstances whether it is a predictive success and it depends on one's epistemological beliefs whether or not one speaks of an explanatory success.

Now, it turns out to be very illuminating to write out in detail what is implicitly well-known from Hempel's and Popper's work, viz., that the HD method applied to theories is essentially a stratified, two-step method, based on a macro- and a micro-argument, with much room for complications. In the macro-step already indicated, one derives GTI's from the theory. In their turn, such GTI's are tested by deriving from them, in the micro-step, with the help of suitable initial conditions, testable individual statements, called Individual Test Implications (ITI's). The suggested decomposition amounts in some detail to the following.

For the macro-argument we get:

Theory: X	
Logico-Mathematical Claim (LMC): if X then I	
	Modus Ponens (MP)
<hr/>	
General Test Implication (GTI): I	

A GTI is assumed formally to be of the form:

I : for all x in D [if $C(x)$ then $F(x)$]

that is, for all x in the domain D , satisfying the initial conditions $C(x)$, the fact $F(x)$ is 'predicted'. All specific claims about x are supposed to be formulated in observation terms.

Successive testing of a particular GTI I will lead to one of two mutually exclusive results. The one possibility is that sooner or later we get falsification of I by coming across a falsifying instance or counterexample of I . Although a counterexample of I is, strictly speaking, also a counterexample of X , I also call it, less dramatically, a *negative instance* of or an *individual problem* for X . The alternative possibility is that, despite variations in members of D and ways in which C can be satisfied, all our attempts to falsify I fail, i.e., lead to the predicted results. The conclusion attached to repeated success of I is of course that I is established as true, i.e., as a general (reproducible) fact. I will call such an I a (*general*) *success* of X . Finally, it may well be that certain GTI's of X have already been tested long before X was taken into consideration. The corresponding individual problems and general successes have to be included in the evaluation report of X (see below).

Recorded problems and successes are (partial) answers to the success question: what are the potential successes and problems of the theory? Hence, testing GTI's derived in accordance with the macro HD argument is effective in answering this question. Moreover, it is efficient, for it will never lead to irrelevant, neutral results, that is, results that are neither

predicted by the theory nor in conflict with it. Neutral results for one theory only come into the picture when we take test results of other theories into consideration, that is, the comparative evaluation of two or more theories (see the next section).

I call the list of partial answers to the success question, which are available at a certain moment t , the *evaluation report* of X at t , consisting of the following two components:

- the set of individual problems,*
i.e., established counterexamples of GTI's of X ,
- the set of general successes,*
i.e., the established GTI's of X , that is, general facts derivable from X .

Hence, the goal of separate theory evaluation can be explicated as aiming at such an evaluation report.

Models of HD evaluation

Let us now have a closer look at the testing of a general test implication, the micro-step of the HD method, or, more generally, the testing of a *General Testable Conditional* (GTC). The micro HD argument amounts to:

General Test Conditional (GTC): G : for all x in D [if $C(x)$ then $F(x)$]	
Relevance Condition: a in D	
<hr/>	
Individual Test Conditional: if $C(a)$ then $F(a)$	Universal Instantiation (UI)
Initial Condition(s) (IC): $C(a)$	
<hr/>	
Individual Test Implication (ITI): $F(a)$	Modus Ponens (MP)

If the specific prediction posed by the individual test implication turns out to be false, then the hypothesis G has been falsified. The relevant (description of the) object may be called a *counterexample* or a *negative instance* or an *individual problem* of G . If the specific prediction turns out to be true the relevant (description of) the object may be called a *positive instance* or an *individual success* of G . Besides positive and negative instances of G , we may want to speak of *neutral instances* or *neutral results*. They will not arise from testing G , but they may arise from testing other general test implications.

Consequently, the evaluation report of GTC's basically has two sides, like the evaluation reports of theories; one for problems and the other for successes. Again, they form partial answers to the success question now raised by the GTC. However, here the two sides list entities of the same kind: negative or positive instances, that is, individual problems and individual successes, respectively.

It is again clear that the micro HD argument for a GTC G is effective and efficient for making its evaluation report: each test of G either leads to a positive instance, and hence to an increase of G 's individual successes, or it leads to a negative instance, and hence to an increase of G 's individual problems. It does not result in neutral instances. Note that what I have described above is the micro HD argument for *evaluating* a GTC. When we confine our attention to establishing its truth-value, and hence stop with the first counterexample, it is the (micro) HD argument for *testing* the GTC.

Concatenation of the macro and micro HD argument gives the full argument for theory evaluation leading to individual problems and individual successes. Instead of the two-step concatenated account, theory evaluation can also be presented completely in terms of contracted HD evaluation, without the intermediate GTI's, leading directly to individual problems and individual successes.

Any application of the HD method (concatenated or contracted) leading to an evaluation report with individual problems and individual successes will be called an application of the *micro-model* of HD evaluation. It is clear that application of the micro-

model is possible for all kinds of general hypotheses, from GTC's to theories with proper theoretical terms.

However, as far as theories which are not just GTC's are concerned the macro-step also suggests the model of *asymmetric* HD evaluation of a theory, leading to an evaluation report with individual problems and general successes. In that case, GTI's are derived in the macro-step, and only tested, not evaluated, in the micro-step. In the micro-model of HD evaluation of theories, in particular when contraction is used, the intermediate general successes of theories may disappear from the picture. However, in scientific practice, these intermediate results frequently play an important role. The individual successes of theories are summarized, as far as possible, in general successes. These general successes relativize the dramatic role of falsification via other general test implications. As we shall see in the next section, they form a natural unit of merit for theory comparison, together with counterexamples, as the unit of (individual) problems. In the next section, the model of asymmetric HD evaluation plays a dominant role. The results it reports will then be called counterexamples and (general) successes.

However, individual problems frequently can be summarized in terms of 'general problems'. They amount to established 'falsifying general hypotheses' in the sense of Popper. Hence, there is also room for a *macro-model* of HD evaluation, where, besides general successes, the evaluation report lists general problems as well. In this case, all individual successes and individual problems are left out of the picture as long as they do not fit into an established general success or problem. Note that there is also the possibility of a fourth model of HD evaluation of an asymmetric nature, with individual successes and general problems, but as far as I can see, it does not play a role in scientific practice.

The three interesting models of HD evaluation of theories can be ordered in terms of increasing refinement: the macro-model, the asymmetric model, and the micro-model.

It can be shown that the main lines of the analysis of testing and evaluation also apply when the test implications are of a statistical nature. However, for deterministic test implications there are already all kinds of complications of testing and evaluation, giving occasion to 'dogmatic strategies' and suggesting a refined scheme of HD argumentation. Although such problems multiply when statistical test implications are concerned, I shall restrict myself to a brief indication of those in the deterministic case.

Complicating factors

According to the idealized versions of HD testing and evaluation presented so far there are only cases of evident success or failure. However, as is well known, several factors complicate the application of the HD method. Let us approach them first from the falsificationist perspective. Given the fact that scientists frequently believe that their favorite theory is (approximately) true, they have, on the basis of these factors, developed strategies to avoid the conclusion of falsification. The important point of these dogmatic or conservative strategies is that they may rightly save the theory from falsification, because the relevant factor may really be the cause of the seeming falsification. Although the recognition of a problem for a theory is more dramatic from the falsificationist perspective, when evaluating a theory one may also have good reasons for trying to avoid a problem.

I distinguish five complicating factors, each leading to a standard saving strategy. They show in detail, among other things, how Lakatos' methodology of research programs (Lakatos, 1970/1978), saving the hard core, can be defended and effected. Though perhaps less frequently practiced, the same factors may also be used, rightly or wrongly, as point of impact for contesting some success. In this case, there is even one additional factor. All six factors concern suppositions in the concatenated macro and micro HD argument. I do not claim originality with these factors as such; most of them have been mentioned by Lakatos and have been anticipated by Hempel, Popper and others. However, their subsequent systematic survey and localization is made possible by the decomposition of the macro and micro HD argument. It is left to the reader to identify examples of the factors.

In the subjoined, refined schematization of the concatenated HD arguments the six main vulnerable factors or weak spots in the argument have been made explicit and

emphasized by a 'Q', indicating 'Questionable'. The relevant assumptions have been given suggestive names, such that they may be assumed to be self-explanatory. Some of them have been grouped together by the numbering because of their analogous logical role.

Theory: X	
Q1.1: Auxiliary hypotheses: A	
Q1.2: Background Knowledge: B	
Q2: Logico-Mathematical Claim (LMC): if X, A, B then I	
<hr/>	
Modus Ponens (MP)	
General Test Implication (GTI): I : for all x in D [if $C(x)$ then $F(x)$]	
Q3: Observation presuppositions: $C=C^*, F=F^*$	
<hr/>	
General Test Implication (GTI*): I : for all x in D [if $C^*(x)$ then $F^*(x)$]	
Q4.1: Relevance Condition: a in D	
<hr/>	
Universal Instantiation (UI)	
Individual Test Conditional: if $C^*(a)$ then $F^*(a)$	
Q4.2: Initial Condition(s) (IC): $C^*(a)$	
<hr/>	
Modus Ponens (MP)	
Individual Test Implication (ITI): $F^*(a)$	
Data from repeated tests	
Q5: <hr/>	
Decision Criteria	
<hr/>	
<i>either</i> sooner or later a counterexample of GTI*, leading to the conclusion not-GTI*	<i>or</i> only positive instances of GTI*, suggesting inference of GTI* by Inductive Generalization Q6

The consequence of the first five factors (auxiliary hypotheses + background knowledge claims, logico-mathematical claims, observation presuppositions, initial conditions, and decision criteria) is that a negative outcome of a test of a theory only points unambiguously in the direction of falsification under certain conditions. Falsification of the theory only follows when it may be assumed that the auxiliary hypotheses, the background knowledge claims and the observation presuppositions are (approximately) true, that the logico-mathematical claim is valid, that the initial conditions were indeed realized and that the used decision criteria were adequate in the particular case. Hence, it will not be too difficult to protect a beloved theory from threatening falsification by challenging one or more of these suppositions.

If the truth question regarding a certain theory is the guiding question, most points of this section, e.g., the decomposition of the HD method, the evaluation report and the survey of complications, are only interesting as long as the theory has not been falsified. However, if one is also, or primarily, interested in the success question the results remain interesting after falsification. In the next section I will show how this kind of separate HD evaluation can be put to work in comparing the success of theories. Among other things, this application explains and even justifies non-falsificationist behavior, including certain kinds of dogmatic behavior.

6. Empirical Progress and Pseudoscience

The analysis of separate HD evaluation has important consequences for theory comparison and theory selection. The momentary evaluation report of a theory immediately suggests a plausible way of comparing the success of different theories. Moreover, it suggests the further testing of the comparative hypothesis that a more successful theory will remain more successful and, finally, the rule of theory selection, prescribing its adoption, for the time being, if it has so far proven to be more successful. The suggested comparison and rule of selection will be based on the asymmetric model of evaluation in terms of general successes

and individual problems. However, it will also be shown that the symmetric approach, in terms of either individual or general successes and problems, leads to an illuminating symmetric evaluation matrix, with corresponding rules of selection.

Asymmetric theory comparison

A central question for methodology is what makes a new theory better than an old one. The intuitive answer for the new theory being as good as the old is plausible enough. The new theory has at least to save the established strengths of the old one and not to add new weaknesses on the basis of the former tests. In principle, we can choose any combination of individual or general successes and problems to measure strengths and weaknesses. However, the combination of general successes and individual problems, i.e., the two results of the asymmetric model of (separate) HD evaluation, is the most attractive. First, this combination seems the closest to actual practice and, second, it turns out to be the most suitable one for a direct link with questions of truth approximation. For these reasons I will first deal with this alternative and come back to the two symmetric alternatives.

Given the present choice, the following definition is the obvious formal interpretation of the idea of (*prima facie*) progress, i.e., *increasing success*:

Theory *Y* is (at time *t*) *at least as successful as (more successful than or better than)* theory *X* iff (at *t*)

- all individual problems of *Y* are (individual) problems of *X*
- all general successes of *X* are (general) successes of *Y*
- (- *Y* has extra general successes or *X* has extra individual problems)

The definition presupposes, of course, that for every recorded (individual) problem of one theory, it has been ascertained whether or not it is also a problem for the other, and similarly whether or not a (general) success of one is also a success of the other. The first clause may be called the 'instantial clause' as appealing and relatively neutral. From the realist perspective it is plausible to call the second clause the 'explanatory clause'. From other epistemological perspectives one may choose another, perhaps more neutral name, such as, the general success clause. It is also obvious how one should define, in similar terms to those above, the general notion of 'the most successful theory thus far among the available alternatives' or, simply, 'the best (available) theory'.

It should be stressed that the diagnosis that *Y* is more successful than *X* does not guarantee that this will remain the case. It is a *prima facie* diagnosis based only on facts established thus far, and new evidence may change the comparative judgment. But, assuming that established facts are not called into question, it is easy to check that the judgement cannot have to be reversed, i.e., that *X* becomes more successful than *Y* in the light of old and new evidence. For, whatever happens, *X* has extra individual problems or *Y* has extra general successes.

It should be conceded that it will frequently not be possible to establish the comparative claim, let alone that one theory is more successful than all its available alternatives. The reason is that these definitions do not guarantee a constant linear ordering, but only an evidence-dependent partial ordering of the relevant theories. In other words, in many cases there will be 'divided success': one theory has successes another does not have, and *vice versa*, and similarly for problems. Of course, one may interpret this as a challenge for refinements, e.g., by introducing different concepts of 'relatively maximal' successful theories or by a quantitative approach. However, it will become clear that in case of 'divided success' another heuristic-methodological approach, of a qualitative nature, is more plausible.

As a matter of fact, the core of HD evaluation amounts to several heuristic principles. The first principle says that, as long as there is no best theory, one may continue the separate HD evaluation of all available theories. The aim is, of course, to explore the domain further in terms of general facts to be accounted for and individual problems to be overcome by an overall better theory. For the moment, I will concentrate on the second principle, applicable in

the relatively rare case that one theory is more successful than another one, and hence in the case that one theory is the best.

Suppose theory *Y* is at *t* more successful than theory *X*. This condition is not yet a sufficient reason to prefer *Y* in some substantial sense. That would be a case of 'instant rationality'. However, when *Y* is at a certain moment more successful than *X*, this situation suggests the following comparative success hypothesis:

CSH: *Y* (is and) will remain more successful than *X*

CSH is an interesting hypothesis, even if *Y* is already falsified. Apart from the fact that *Y* is known to have some extra successes or *X* some extra individual problems at *t*, CSH amounts at *t* to two components, one about problems, and the other about successes:

CSH-P: all individual problems of *Y* are individual problems of *X*

CSH-S: all general successes of *X* are general successes of *Y*

where 'all' is to be read as 'all past and future'.

Although there may occasionally be restrictions of a fundamental or practical nature, these two components concern, in principle, testable generalizations. Hence, testing CSH requires application of the micro HD argument. Following CSH-P, we may derive a GTI from *Y* that does not follow from *X*, and test it. When we get a counterexample of this GTI, and hence an individual problem of *Y*, it may be ascertained if the problem is shared by *X*. If it is not, we have falsified CSH-P.

Alternatively, following CSH-S, we may derive a GTI from *X* which cannot be derived from *Y*, and test it. If it becomes accepted, its acceptance means falsification of CSH-S. Of course, in both cases, the opposite test result confirms the corresponding comparative subhypothesis, and hence CSH, and hence increases the registered success difference. In the following, for obvious reasons, I call (these two ways of) testing CSH *comparative HD evaluation*.

The plausible rule of theory selection is now the following:

Rule of Success (RS)

When *Y* has so far proven to be more successful than *X*, i.e., when CSH has been 'sufficiently confirmed' to be accepted as true, eliminate *X* in favor of *Y*, at least for the time being.

RS does not speak of 'remaining more successful', for that would imply the presupposition that the CSH could be completely verified (when true). Hence I speak of 'so far proven to be more successful' in the sense that CSH has been 'sufficiently confirmed' to be accepted as true; that is, CSH is accepted as a (twofold) inductive generalization. The point at which CSH is 'sufficiently confirmed' will be a matter of dispute. Be this as it may, the acceptance of CSH and consequent application of RS is the core idea of *empirical progress*, a new theory that is better than an old one. RS may even be considered as the (fallible) criterion and hallmark of scientific rationality, acceptable for the empiricist as well as for the realist.

As soon as CSH is (supposed to be) true, the relevance of further comparative HD evaluation is diminished. Applying RS, i.e., selecting the more successful theory, then means the following, whether or not that theory already has individual problems. One may concentrate on the further separate HD evaluation of the selected theory, or one may concentrate on the attempt to invent new interesting competitors, that is, competitors that are at least as successful as the selected one.

Given the tension between reducing the set of individual problems of a theory and increasing its (general observational) successes, it is not an easy task to find such interesting competitors. The search for such competitors cannot, of course, be guided by prescriptive rules, like RS, but there certainly are heuristic principles of which it is easy to see that they stimulate new applications of RS. Let me start by explicitly stating the two suggested

principles leading to RS. First, there is the *principle of separate HD evaluation* (PSE): "Aim via general test implications to establish new laws which can be derived from your theory (general successes) or, equivalently, aim at new negative instances (individual problems) of your theory". Secondly, the *principle of comparative HD evaluation* (PCE): "Aim at HD testing of the comparative success hypothesis, when that hypothesis has not yet been convincingly falsified". In both cases, a typical Popperian aspect is that one should aim at deriving test implications, which are, in the light of the background knowledge, very unlikely or even impossible. The reason is, of course, that a (differential) success of this kind is more impressive than that of a more likely test implication. In view of the first comparative (confirmation) principle (P1), such a success leads in case of PSE to more confirmation of a theory, assuming that that has not yet been falsified, and in case of PCE to more confirmation of the comparative success hypothesis in general.

As already suggested, RS presupposes previous application of PSE and PCE. But some additional heuristic principles, though not necessary, may also promote the application of RS. To begin with, the *principle of content* (PC) may do so: "Aim at success preserving, strengthening or, *pace* Popper, weakening your theory". A stronger theory is likely to introduce new individual problems but gain new general successes. If the latter arise and the former do not materialize, RS can be applied. Something similar applies to a weaker theory. It may solve problems without sacrificing successes. I would also like to mention the *principle of dialectics* (PD) for two theories that escape RS because of divided success: "Aim at a success preserving synthesis of two RS-escaping theories". In ICR (Section 8.3), I explicate a number of dialectical notions in this direction. Of course, there may come a point at which further attempts to improve a theory and hence to discover new applications of RS are abandoned.

In sum, the asymmetric model of HD evaluation of theories naturally suggests the definition of 'more successful', the comparative success hypothesis, the testing of such a hypothesis, i.e., comparative HD evaluation, and the rule of success (RS) as the cornerstone of empirical progress. Separate and comparative HD evaluation provide the right ingredients for applying first the definition of 'more successful' and, after sufficient tests, that of RS, respectively. In short, separate and comparative HD evaluation are functional for RS, and HD testing evidently is functional for both types of HD evaluation. The method of HD evaluation of theories combined with RS and the principles stimulating the application of RS might well be called the instrumentalist methodology. In particular, it may be seen as a free interpretation or explication of Laudan's problem solving model (Laudan, 1977), which is generally conceived as a paradigm specification of the idea of an instrumentalist methodology. However, it will also be called, more neutrally, the *evaluation methodology*. It will be said that RS governs this methodology. The claim is that this methodology governs the short-term dynamic of science, more specifically, the internal and competitive development of research programs.

Note that the evaluation methodology demonstrates continued interest in a falsified theory. The reasons behind it are easy to conceive. First, it is perfectly possible that the theory nevertheless passes other general test implications, leading to the establishment of new general successes. Second, even new tests leading to new individual problems are very useful, because they have to be overcome by a new theory. Hence, at least as long as no better theory has been invented, it remains useful to evaluate the old theory further in order to reach a better understanding of its strengths and weaknesses.

Symmetric theory comparison

The symmetric models of separate HD evaluation, i.e., the micro- and the macro-models, suggest a somewhat different approach to theory comparison. Although these approaches do not seem to be in use to the extent of the asymmetric one and can only indirectly be related to truth approximation, they lead to a very illuminating (comparative) evaluation matrix.

A better theory has to be at least as successful as the old one, and this fact suggests general conditions of adequacy for the definitions of a 'success', of a 'problem' and of a 'neutral result'. The asymmetric definition of 'at least as successful' presented above only deals

explicitly with individual problems and general successes; neutral results remain hidden, but it is easy to check that they nevertheless play a role. The symmetric models take all three types of results explicitly into account. The macro-model focuses on such results of a general nature, the micro-model on such results of an individual nature.

The notions of general successes and general problems are not problematic. Moreover, general facts are neutral for a theory when they are neither a problem nor a success. A better theory retains general successes as (already tested) general test implications, and does not give rise to new general test implications of which testing leads to the establishment of new general problems. Moreover, general problems may be transformed into neutral facts or even successes, and neutral general facts may be transformed into successes. The notions of individual successes, individual problems and neutral results are not problematic either, as long as we list them in terms of positive, negative and neutral instances, respectively. A better theory keeps the positive instances as such; it does not lead to new negative instances, and neutral instances may remain neutral or become positive. However, if we want to list individual successes and/or individual problems in terms of statements, the situation becomes more complicated, but it is possible (see ICR, 116-7).

Let us now look more specifically at the symmetric micro-model, counting in terms of individual problems, successes and neutral results, that is, negative, positive and neutral instances or (statements of) individual facts. Hence, in total, the two theories produce a matrix of nine combinations of possible instances or individual facts. In order that the matrix can also be made useful for the macro-model, I present it in terms of facts. For the moment, these facts are to be interpreted as individual facts. The entries represent the status of a fact with respect to the indicated theories *X* and *Y*.

		<i>X</i>		
		Negative	neutral	Positive
<i>Y</i>	negative	B4: 0	B2: –	B1: –
	neutral	B8: +	B5: 0	B3: –
	positive	B9: +	B7: +	B6: 0

The (comparative) evaluation matrix

From the perspective of *Y* the boxes B1/B2/B3 represent unfavorable facts (indicated by '–'), B4/B5/B6 (comparatively neutral or) indifferent facts (0), and B7/B8/B9 favorable facts (+). The numbering of the boxes, anticipating a possible quantitative use, was determined by three considerations: increasing number for increasingly favorable results for *Y*, a plausible form of symmetry with respect to the diagonal of indifferent facts, and increasing number for indifferent facts that are increasingly positive for both theories.

It is now highly plausible to define the idea that *Y* is *more successful* than *X* in the light of the available facts as follows: there are no unfavorable facts and there are some favorable facts, that is, B1/2/3 should be empty, and at least one of B7/8/9 non-empty. This state of affairs immediately suggests modified versions of the comparative success hypothesis and the rule of success.

It is also clear that, by replacing individual facts by general facts, we obtain macro-versions of the matrix, the notion of comparative success, the comparative success hypothesis and the rule of success. A general fact may be a general success, a general problem or a neutral general fact for a theory.

In all these variants, the situation of being more successful will again be rare, but it is certainly not excluded. In ICR (Chapter 11) I argue, for instance, that the theories of the atom developed by Rutherford, Bohr and Sommerfeld can be ordered in terms of general facts according to the symmetric definition.

		Light propagation experiments						Experiments from other fields						
Theories	Experimental facts	Aberration	Fizeau convection coefficient	Michelson-Morley	Kennedy-Thorndike	Moving sources and mirrors	De Sitter spectroscopic binaries	Michelson-Morley, using sunlight	Variation of mass with velocity	General mass-energy equivalence	Radiation from moving charges	Meson decay at high velocity	Trouton-Noble	Unipolar induction
Ether theories	Stationary ether, no contraction	A	A	D	D	A	A	D	D	N	A	N	D	D
	Stationary ether, Lorentz contraction	A	A	A	D	A	A	A	A	N	A	N	A	D
	Ether attached to ponderable bodies	D	D	A	A	A	A	A	D	N	N	N	A	N
Emission theories	Original source	A	A	A	A	A	D	D	N	N	D	N	N	N
	Ballistic	A	N	A	A	D	D	D	N	N	D	N	N	N
	New source	A	N	A	A	D	D	A	N	N	D	N	N	N
Special theory of relativity		A	A	A	A	A	A	A	A	A	A	A	A	A

Comparison of experimental record of seven electrodynamic theories.

Legend: A: agreement, D: disagreement, N: not applicable

Another set of examples of this kind is provided by the table (adapted from: Panofsky and Phillips, 1962², p. 282), representing the records in the face of 13 general experimental facts of the special theory of relativity (STR) and six alternative electrodynamic theories, viz., three versions of the ether theory and three emission theories. According to this table, STR is more successful than any of the other; in fact it is maximally successful as far as the 13 experimental facts are concerned. Moreover, Lorentz's contraction version of the (stationary) ether theory is more successful than the contractionless version. Similarly, the ballistic version of the emission theory is more successful than the other two. However, it is also clear that many combinations lead to divided results. For instance, Lorentz's theory is more successful in certain respects (e.g., De Sitter's spectroscopic binaries) than the ballistic theory, but less successful in other respects (e.g., the Kennedy-Thorndike experiments).

In the present approach it is plausible to define, in general, one type of divided success as a liberal version of more successfulness. *Y* is *almost more successful* than *X* if, besides some favorable facts and (possibly) some indifferent facts, there are some unfavorable facts, but only of the B3-type, provided there are (favorable) B8- or B9-facts or the number of B3-facts is (much) smaller than that of their antipodes, that is, B7-facts. The provision clause guarantees that it remains an asymmetric relation. Crucial is the special treatment of B3-facts. They correspond to what is called *Kuhn-loss*: the new theory seems no longer to retain a success demonstrated by the old one. The idea behind their suggested relatively undramatic nature is the belief that further investigation may show that and how a B3-fact turns out to be a success after all, perhaps by adding an additional (non-problematic) hypothesis. In this case it becomes an (indifferent) B6-fact. Hence, the presence of B3-facts is first of all an invitation to further research. If this is unsuccessful, such a B3-fact becomes a case of recognized Kuhn-loss. Unfortunately, the table above does not contain an example of an almost more successful theory.

Cases of divided success may also be approached by some (quasi-)quantitative weighing of facts. Something like the following quantitative evaluation matrix is directly suggested by the same considerations that governed the number ordering of the boxes.

		X		
		negative	neutral	positive
Y	Negative	B4: -1/-1	B2: -3/+3	B1: -4/+4
	Neutral	B8: +3/-3	B5: 0/0	B3: -2/+2
	Positive	B9: +4/-4	B7: +2/-2	B6: +1/+1

The quantitative (comparative) evaluation matrix

All qualitative success orderings of electrodynamic theories to which the table gives rise, remain intact on the basis of this quantitative matrix (which is not automatically the case). Moreover, we now of course get a linear ordering, with Lorentz's theory in the second position after STR and far ahead of the other alternatives. Of course, one may further refine such orderings by assigning different basic weights to the different facts, to be multiplied by the relative weights specified in the quantitative matrix.

Like a similar observation in the symmetric case, it is now possible to interpret the qualitative and the quantitative versions of the evaluation matrix as explications of some core aspects of Laudan's (1977) problem-solving model of scientific progress, at least as far as empirical problems and their solutions are concerned.

Scientific and pseudoscientific dogmatism

Although the method of HD testing, HD evaluation, and hence the evaluation methodology have a falsificationist flavor, each with its own aim, they are certainly not naive in the sense in which Popper's methodology has sometimes been construed. Naive falsificationism in the sense described by Lakatos (1970/1978) roughly amounts to applying HD testing for purposes of theory evaluation and elimination. Its core feature then becomes to further discard (convincingly) falsified theories. Lakatos has also construed a sophisticated version of falsificationism such that, when comparing theories, he takes their 'unrefuted content' into account, a practice that allows falsified theories to remain in the game. Moreover, Lakatos has proposed a 'methodology of research programs', which operates in a sophisticated falsificationist way. However, it works in such a way that it postpones the recognition of falsifications of the 'hard core theory' as long as it is possible to roll off the causes of falsification dogmatically onto auxiliary hypotheses or background theories.

It can be argued that HD evaluation can be seen as an explication of sophisticated falsificationism, leaving room for a dogmatic research program specification. Moreover, it can be argued that the falsificationist and the evaluation methodology may be functional for truth approximation, and that the latter non-falsificationist methodology, ironically enough, is much more efficient for that purpose.

The (naive) falsificationist methodology amounts to restricting the rule of success (RS) to not-yet-falsified theories in combination with the following rule:

Rule of Elimination (RE)

When a theory has been convincingly falsified, elimination should follow, and one should look for a new theory

The evaluation methodology can be summarized by the

Principle of Improvement (of theories) (PI)

Aim at a more successful theory, and successive application of RS

Both methodologies presuppose the

Principle of (Falsifiability or) Testability (PT)

Aim at theories that can be tested, and hence evaluated, in the sense that test implications can be derived, which can be tested for their truth-value by way of observation

Hence, the relativization of the methodological role of falsification, inherent in the evaluation methodology, should not be construed as a plea to drop falsifiability as a criterion for being an empirical theory. On the contrary, empirical theories are supposed to be able to score successes or, to be precise, general successes. Moreover, PI presupposes the principles of separate and comparative HD evaluation (PSE/PCE) as introduced in Section 5, whereas PE presupposes them only for not yet falsified theories. Finally, it is possible to extend PI to nonempirical features, for example aesthetic features such as simplicity and symmetry. In Sections 8 and 9 I will formally take such features into account in relation to truth approximation. However, it should be clear that their methodological role in theory choice is primarily or even exclusively restricted to cases of equal empirical success (see Kuipers, 2002, Section 6, for a detailed treatment of their role).

It is clear that RE may retard empirical progress in the sense of PI. Moreover, it can also be argued that RE affects the prospects for truth approximation. A striking feature of PI in this respect is that the question of whether the more successful theory is false or not does not play a role at all. That is, the more successful theory may well be false, provided all its counterexamples are also counterexamples of the old theory.

These claims about truth approximation have important methodological consequences. They enable a new explanation, even justification, of the observation of Kuhn, Lakatos and others that there is quite a discrepancy between falsificationist (methodological) theory and non-falsificationist practice. In principle, even with respect to the paradigmatic non-falsificationist method of idealization and concretization, as propagated by Nowak, but this requires 'refined' truth approximation (see Section 10). Straightforward (basic or refined) truth approximation may be seen as the primary, conscious or unconscious, motive for non-falsificationist behavior. Dogmatic behavior, in the sense of working within a research program, is only a secondary motive for non-falsificationist behavior. Whatever the main motive, as long as such behavior is directed at theory improvement within the program, it can be distinguished from pseudoscientific behavior.

The following principle expresses the core idea:

Principle of improvement guided by research programs (PIRP)

One should primarily aim at progress within a research program, i.e., aim at a better theory while keeping the hard core of the program in tact. If, and only if, this strategy does not work, try to adapt the hard core, while leaving the vocabulary in tact. If, and only if, this second strategy is also unsuccessful, look for another program with better perspectives on progress

Whereas responsible dogmatic behavior is governed by this refined principle of improvement, leaving room for dogmas, one of the typical marks of pseudoscientific behavior is that one is usually not even aiming at improvement by the first strategy, let alone by the second.

Our notion of comparative evaluation is governed by the notion of being '(almost) more successful'. This is a rather strict strategy. In ICR I question the general usefulness of quantitative liberalizations of 'successfulness', and for that matter, of 'truthlikeness', mainly because they need real-valued distances between models, a requirement which is very unrealistic in most scientific contexts. Hence, the applicability of liberal notions may well be laden with arbitrariness. Be this as it may, it is important to stress that the strict strategy does not lead to void or almost void methodological principles. If there is divided success between theories, the Principle of Improvement amounts, more specifically, to the already mentioned recommendation that we should try to apply the Principle of Dialectics: "Aim at a success preserving synthesis of the two RS-escaping theories", of course, with a plausible program-bound version. Hence, the restricted applicability of the strict notion of comparative success

does not exclude the possibility of clear challenges being formulated in cases where they do not apply – – on the contrary.

III BASIC TRUTH APPROXIMATION

This part introduces and analyzes the theory of naive or basic truth approximation and its relation to empirical progress and confirmation, first for epistemologically unstratified theories and later for stratified ones.

In Section 7 the qualitative idea of truthlikeness is introduced, more specifically the idea that one theory can be closer to the truth than another, which is called 'nomic truthlikeness'. Here 'the truth' concerns 'the nomic truth', i.e., the strongest true hypothesis, assumed to exist according to the 'nomic postulate', about the physical or nomic possibilities, called 'the nomic world', restricted to a given domain and, again, as far as can be expressed within a given vocabulary. The Success Theorem is crucial, according to which 'closer to the truth' implies 'being at least as successful', even straightforwardly, if defined in the asymmetric way. It is used to argue that the evaluation methodology is effective and efficient for nomic truth approximation.

ICR Ch. 7, moreover, deals with 'actual' truthlikeness and truth approximation, where 'the actual truth' represents the actual possibility or (restricted) world, or their historical succession. The chapter results in a survey of bifurcations of truthlikeness theories and concludes with their plausible methodological and epistemological consequences for the notions of novel facts, crucial experiments, inference to the best explanation and descriptive research programs.

Section 8 argues that 'basic' nomic truthlikeness and the corresponding methodology have plausible conceptual foundations, of which the dual foundation will be the most appealing to scientific common sense: 'more truthlike' amounts to 'more true consequences and more correct models', in line with the asymmetric definition of 'more successful'. There is also an indication of how this analysis leaves room for nonempirical considerations in theory evaluation, such as aesthetic ones.

In ICR Ch. 8 a detailed comparison is presented between Popper's original definition of truthlikeness and the basic definition, showing among other things that the latter does not have the generally recognized shortcomings of the former. Moreover, it is also argued that basic truthlikeness suggests a non-standard, viz., intralevel rather than interlevel, explication of the main intuitions governing the so-called correspondence theory of truth. Moreover, it is made clear that the presented cognitive structures suggest logical, methodological and ontological explications of some main dialectical concepts, viz., dialectical negation, double negation, and the triad of thesis-antithesis-synthesis.

Section 9 introduces the first major sophistication, the stratification arising from the (changing) distinction between observational and theoretical terms, leading to the distinction between observational and theoretical truth approximation.

In ICR Ch. 9 this also leads to the idea of 'the referential truth', i.e., the truth about which terms of a vocabulary refer and which do not, where 'reference' gets a precise definition on the basis of the nomic postulate. This enables the definition of the referential claim of a theory and hence of the idea of one theory being closer to the referential truth than another. Moreover, the overall analysis is shown to lead to plausible rules of inference to the best theory, viz. as the closest to the observational, the theoretical, and the referential truth.

For readers with a model theoretic background it is important to realize the main divergence between the (dominant) model theoretic view on empirical theories and my favorite so-called 'structuralist' perspective on them. According to the former the target of theorizing is one particular 'intended application', the actual world, and according to the latter it is a set of 'intended applications', the nomic possibilities. Although the suggested model theoretic perspective may be dominant, I would like to leave room for the possibility that an alternative model theory in line with the structuralist perspective will be further developed and become respected. Not in order to replace the dominant one but in order to obtain an

alternative that is more suitable for certain purposes. However, I do not see this alternative as a non-Tarskian move in some deep sense. Starting from Tarski's basic definition, which is that of 'truth in a structure' (Hodges, 1986), and assuming that one has more than one intended application in mind, it is plausible to define that a theory is true if and only if it is true for all intended applications. However, in this case there are at least two coherent ways of defining that a theory is false. In line with the standard approach one may be inclined to call a theory only false when it is false for all intended applications, and indeterminate if it is neither true for all intended applications nor false. However, it is in line with the structuralist approach to call a theory already false if it is false for at least one intended application.

7. Truthlikeness and Truth Approximation

I shall first deal with the logical or conceptual problem of defining '(more) truthlikeness', assuming that we know what 'the truth' is. I then turn to the prospects for truth approximation by using the method of HD evaluation. In this section we do not yet assume a distinction between theoretical and observational terms, which amounts to assuming that all terms are observational.

Truthlikeness

The starting point of the idea of truthlikeness is a vocabulary and a domain. A conceptual possibility is a situation or state of affairs that can be described in the vocabulary, and is therefore conceivable. Let CP be the set of all *conceptual possibilities* that can be described in terms of the vocabulary, also called the *conceptual frame*. A theory will be associated with a subset of CP . A basic assumption, the *Nomic Postulate*, is that the representation of the chosen domain in terms of the vocabulary results in a unique subset of CP containing the *nomic possibilities*. We can identify this usually unknown subset with the truth T for reasons that will become clear shortly. For the sake of convenience I here assume that we can somehow characterize T in terms of the vocabulary.

The aim of theory formation is the actual characterization of T . Hence, the nomic possibilities constituting T can also be called *desired* possibilities, and the elements in $CP - T$, representing the nomic *impossibilities*, can also be called the *undesired* possibilities. A theory X consists of a subset X of CP , with the strong claim " $X = T$ ". If X encloses T , X does not exclude desired possibilities. Thus the weaker claim " $T \subseteq X$ ", meaning that X admits all desired possibilities, is true, in which case we will also say that X is true as a hypothesis. If this weaker claim is false we will also say that X is false as a hypothesis. If $T \subseteq Y \subset X$, Y excludes more undesired possibilities than X and so the claim " $T \subseteq Y$ ", that goes with it, is stronger than " $T \subseteq X$ ", but nevertheless true. In this sense theory T itself is the strongest true theory, and I call it *the truth*. It seems useful to call the elements of X (its) *admitted* possibilities and those of $CP - X$ the *excluded* possibilities (of X). Now it is important to note that the elements of $X \cap T$ are the desired possibilities admitted by X , and $X - T$ consists of the undesired possibilities admitted by X . In Figure 1 all four resulting categories are depicted.

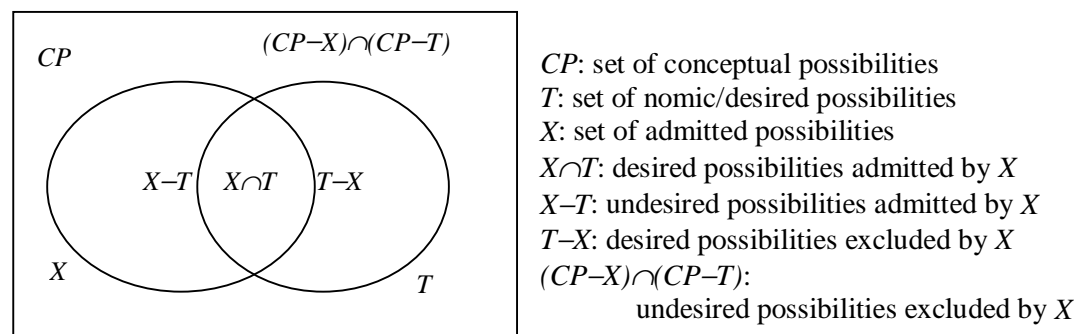


Figure 1: Four categories of possibilities

This brings us directly to the basic definition of (equal or greater) truthlikeness:

Definitions

Y is at least as close to T as X (or: Y resembles T as much as X) iff

(DP) all desired possibilities admitted by *X* are also admitted by *Y*

(UP) all undesired possibilities admitted by *Y* are also admitted by *X*

Y is (two-sided) closer to T than X (or: Y resembles T more than X) iff

(DP) & (DP+) *Y* admits extra desired possibilities

(UP) & (UP+) *X* admits extra undesired possibilities

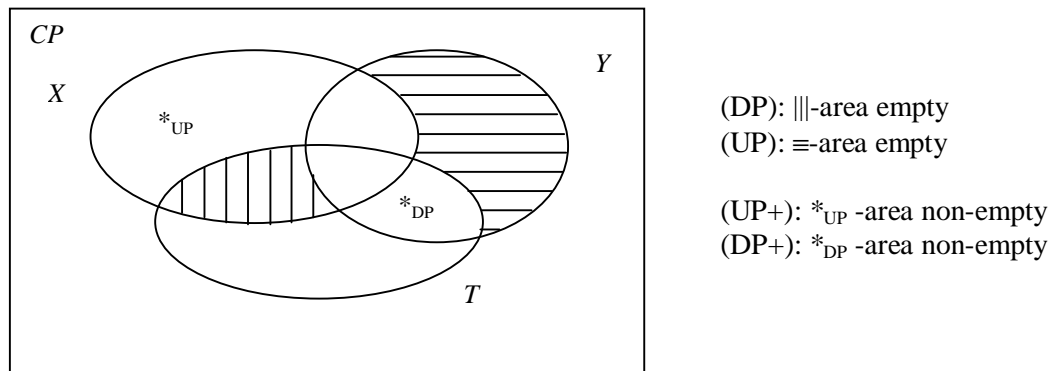


Figure 2: *Y* is closer to the truth *T* than *X*

Figure 2 indicates which sets must be empty (clause (DP) and (UP): vertical and horizontal shading, respectively) and which sets have to be non-empty (clause (DP+) and (UP+): area *DP and area *UP non-empty, respectively) in the case that *Y* is closer to the truth than *X*.

Truth approximation

Now we are able to rephrase the notion of empirical progress and sketch its relation to nomic truth approximation, assuming that *T* is unknown. Recall that as far as theories are concerned, we have dealt up to now with the logical problem of defining nomic truthlikeness, assuming that *T*, the set of nomic possibilities, is at our disposal. In actual scientific practice we don't know *T*; it is the target of our theoretical and experimental efforts. I will now explicate the idea that one theory is more successful than another in terms of 'realized conceptual possibilities' and show that this can be explained by the hypothesis of nomic truth approximation, that is, the hypothesis that the first theory is closer to the truth than the second.

First it is important to note that you can establish that a certain conceptual possibility is nomically possible by experimentally realizing this possibility, but you cannot establish that a certain conceptual possibility is nomically impossible in a direct way, for you cannot realize nomic impossibilities. The standard (partially) indirect way to circumvent this problem is by establishing on the one hand nomic possibilities by realizing them, and on their basis establishing (observational) laws on the other. As we have seen in the preceding section, this is precisely what the separate HD evaluation of theories amounts to. That is, theories are evaluated in terms of their capacity to respect the realized possibilities, i.e., to avoid counterexamples, and to entail the observational laws, i.e., to have general successes.

The problems and successes of a theory will have to be expressed in terms of the data to be accounted for. The data at a certain moment *t* can be represented as follows. Let *R(t)* indicate the set of realized possibilities up to *t*, i.e., the accepted instances (of *T*), which have to be admitted by a theory. Note that there may be more than one realized possibility at the same time, before or at *t*, with plausible restrictions for overlapping domains.

Up to t there will also be some accepted general hypotheses, *the (explicitly) accepted laws*, which have to be accounted for by a theory. On their basis, *the strongest accepted law* to be accounted for is the general hypothesis $S(t)$ associated with the intersection of the sets constituting the accepted laws. It claims that all nomic possibilities satisfy its condition, i.e., it claims that " $T \subseteq S(t)$ ". Of course, $S(t)$ is, via the laws constituting it, in some way or other based on $R(t)$; minimally we may assume that $R(t)$ is not in conflict with $S(t)$, that is, $R(t)$ is a subset of $S(t)$. In the following, however, I shall need the much stronger *correct data (CD-)hypothesis* $R(t) \subseteq T \subseteq S(t)$, guaranteeing that $R(t)$ only contains nomic possibilities, and that hypothesis $S(t)$ only excludes nomic impossibilities. $S(t)$ is thus (assumed to be) true as a hypothesis and may hence rightly be called a law. Henceforth I assume the CD-hypothesis. $R(t)$ may now be called the set of *established nomic possibilities*, and $S(t)$ *the strongest established law*. In fact, for every superset H of $S(t)$ (but subset of CP), hence $S(t) \subseteq H \subseteq CP$, the claim " $T \subseteq H$ " is also true, for which reason H may be called *an (explicitly or implicitly) established law*. Let $Q(S(t))$ indicate the set of all supersets of $S(t)$. Then $Q(S(t))$ represents the set of all *established laws*.

Assuming the data $R(t)$ and $S(t)$, it is now easy to give explications of the notions of individual problems and general successes of a theory X at time t we met in Part II concerning the HD evaluation of theories. The set of *individual problems*, of X at t , is equated with the established members of $R(t) - X$, that is, established nomic possibilities that are not admitted by X . Similarly, the set of *general successes*, of X at t , is equated with the set of established laws that are supersets of X , that is, the members of $Q(S(t)) \cap Q(X)$.

For comparative judgements of the success of theories I shall now explicate the instantial clause of Section 6 in terms of established nomic possibilities, i.e., $R(t)$. Theory Y is *instantially at least as successful as* X if and only if the individual problems of Y form a subset of those of X , that is, Y has no *extra* individual problems. Formally, including some equivalent versions:

$$\begin{aligned} R(t) - Y &\subseteq R(t) - X \\ \Leftrightarrow X \cap R(t) - Y &= \emptyset \Leftrightarrow X \cap R(t) \subseteq Y \cap R(t) \end{aligned}$$

On the other hand, for the explanatory (or general success) clause we have two options for explication, one on the (first) level of subsets of CP and one on the (second) levels of sets of such subsets, leading to two equivalent comparative statements. To begin with the second level, the level of consequences, theory Y is *explanatorily at least as successful as* X if and only if the general successes of X form a subset of those of Y , that is, X has no *extra* general successes. Formally:

$$\begin{aligned} Q(X) \cap Q(S(t)) &\subseteq Q(Y) \cap Q(S(t)) \\ \Leftrightarrow Q(X) \cap Q(S(t)) - Q(Y) &= \emptyset \Leftrightarrow Q(S(t)) - Q(Y) \subseteq Q(S(t)) - Q(X) \end{aligned}$$

On the first level, the level of sets, this is equivalent to the condition that the 'established nomic impossibilities' excluded by X form a subset of those of Y . Formally:

$$\begin{aligned} (CP - S(t)) \cap (CP - X) &\subseteq (CP - S(t)) \cap (CP - Y) \\ \Leftrightarrow Y - X \cup S(t) &= \emptyset \Leftrightarrow Y - S(t) \subseteq X - S(t) \end{aligned}$$

The proof of this equivalence is formally the same as that of the (first) 'equivalence thesis' that will be presented in the next section.

The conjunction of the instantial and the explanatory clause forms the general definition of the statement that one theory is at a certain time at least as successful as another, relative to the data $R(t)/S(t)$. It will be clear that this definition can be seen as an explication of the 'asymmetric' definition given in Section 6. We obtain the strict version, that is, more successful, when in at least one of the two cases proper subsets are concerned. It is called two-sided when proper subsets are involved in both cases. The 'two-sided' strict version is

depicted on the first level in Figure 3 (in which T is indicated by a dotted ellipse to stress that it is unknown).

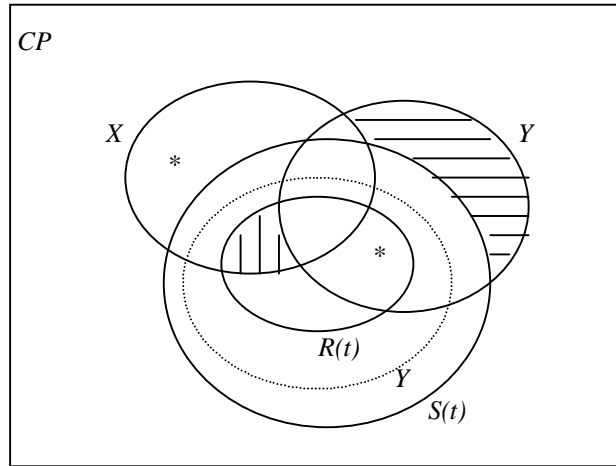


Figure 3: Y is two-sidedly more successful than X relative to $R(t)/S(t)$: shaded areas empty, starred areas non-empty. The unknown T is drawn such that the correct data hypothesis is built in.

Now it is easy to prove the following crucial theorem:

Success Theorem:

If theory Y is at least as close to the nomic truth T as X and if the data are correct then Y (always) remains at least as successful as X .

From this theorem it immediately follows that *success dominance* of Y over X , in the sense that Y is at least as successful as X , can be explained by the following hypotheses: the *truth approximation (TA-)hypothesis*, Y is at least as close to the nomic truth T as X , and the auxiliary *correct data (CD-)hypothesis*.

All notions in the theorem have been explicated, and the proof is, on the first level, only a matter of elementary set-theoretical manipulation, as will be clear from the following presentation of the theorem as an argument:

$Y-T \subseteq X-T$	$T-Y \subseteq T-X$	TA-hypothesis
$T \subseteq S(t)$	$R(t) \subseteq T$	CD-hypothesis
$Y-S(t) \subseteq X-S(t)$	$R(t)-Y \subseteq R(t)-X$	success dominance

As a rule, a new theory will introduce some new individual problems and/or will not include all general successes of the former theory. The idea is that the relative merits can now be explained on the basis of a detailed analysis of the relative 'position' to the truth. However, for such cases a general theorem is obviously not possible.

The importance of The Success Theorem is that it can explain that, and how empirical progress is possible within a conceptual frame CP for a given domain. For this purpose, recall first the Comparative Success Hypothesis (CSH) and the Rule of Success (RS), introduced in Section 6:

CSH: Y (is and) remains more successful than X

RS: When Y has so far been proven to be more successful than X , i.e., when CSH has been 'sufficiently confirmed' to be accepted as true, then eliminate X , in favor of Y , at least for the time being.

For an instrumentalist, CSH and RS are already sufficiently interesting, but the (theory-)realist will only appreciate it for its possible relation to truth approximation, whereas the empiricist and the referential realist will have intermediate interests.

The Success Theorem shows that RS is functional for approaching the truth in the following sense. Assuming correct data, the theorem suggests that the fact that ' Y has so far proven to be more successful than X ' may well be the consequence of the fact that Y is closer to the truth than X . For the theorem enables the attachment of three conclusions to the fact that X has so far proven to be more successful than X ; conclusions which are independent of what exactly 'the nomic truth' is:

- first, it is still possible that Y is closer to the truth than X , a possibility which, when conceived as a hypothesis, the TA-hypothesis, according to the Success Theorem, would explain the greater success in a general way,
- second, it is impossible that Y is further from the truth than X (and hence X closer to the truth than Y), for otherwise, so teaches the Success Theorem, Y could not be more successful,
- third, it is also possible that Y is neither closer nor further from the truth than X , in which case, however, another explanation, now of a specific nature, has to be given for the fact that Y has so far proven to be more successful.

Hence we may conclude that, though 'so far proven to be more successful' does not guarantee that the theory is closer to the truth, it provides good reasons to make this plausible. And this is increasingly the case, the more the number and variation of tests of the comparative success hypothesis increase. It is in this sense that I interpret the claim that RS is functional for truth approximation: the longer the success dominance lasts, despite new experiments, the more plausible that this is the effect of being closer to the truth.

In view of the way in which the evaluation methodology is governed by RS, this methodology is, in general, functional for truth approximation. I would like to spell out this claim in more detail. Recall that the separate and comparative HD evaluation of theories was functional for applying RS in the sense that they precisely provide the ingredients for the application of RS. Recall moreover that HD testing of hypotheses is functional for HD evaluation of theories entailing them. Hence, we get a transitive sequence of functional steps for truth approximation:

- HD testing of hypotheses
 - separate HD evaluation of theories
 - comparative HD evaluation of theories
 - Rule of Success (RS)
 - Truth Approximation (TA)

Consequently, from the point of view of truth approximation, RS can be justified as a prescriptive rule, and HD testing and HD evaluation as its drive mechanisms. Intuitive versions of the rule and the two methods are usually seen as the hallmark of scientific rationality. The analysis of their truth approximating cooperation can be conceived as an explication of what many scientists are inclined to think, and others are inclined to doubt. To be sure, the understanding is not relevant for the practice. That the practice is functional for truth approximation may be conceived as the cunning of reason in science.

It is important to stress once more that RS does not guarantee that the more successful theory is closer to the truth. As long as one does not dispose of explicit knowledge of T , it is impossible to have a rule of success that can guarantee that the more successful theory is

closer to the truth. As we shall see at the end of Section 9, there is only one (near) exception to this claim: purely inductive research.

Another way to summarize the above findings is the following. The TA-hypothesis, claiming that one theory is at least as close to the truth as another, is a perfect example of an empirically testable comparative hypothesis. The Success Theorem says that the TA-hypothesis implies, and hence explains, that the first theory will always be at least as successful as the second.

In terms of an application of HD testing, the Success Theorem amounts to the following claim: the TA-hypothesis has the following two general comparative test implications (assuming the strong, but plausible, auxiliary CD-hypothesis):

all general successes of *X* are general successes of *Y*
all individual problems of *Y* are individual problems of *X*

Note that these are precisely the two components of the comparative success hypothesis (CSH). Hence, when *Y* is at least as successful as *X*, the further HD evaluation, i.e., the further testing of CSH, can indirectly be seen as further HD testing of the TA-hypothesis. When doing so, the latter hypothesis can be falsified, or it can be used to explain newly obtained success dominance.

Recall that I noted in Section 6 that the application of the prescriptive rule RS can be stimulated by several heuristic principles, viz., the *principle of separate HD evaluation* (PSE), the *principle of comparative HD evaluation* (PCE), the *principle of content* (PC), and, finally, the *principle of dialectics* (PD). Of course, we may now conclude that all these principles belonging to the evaluation methodology are indirectly functional for truth approximation.

The Success Theorem is not only attractive from the realist point of view, it is also instructive for weaker epistemological positions, even for the instrumentalist. The Success Theorem implies that a theory that is closer to the truth is also a better derivation instrument for successes, and that the truth (the true theory) is the best derivation instrument. From this not only the self-evident fact follows indirectly that RS and HD evaluation are functional for approaching the best derivation instrument, but also that the heuristic of the realist may also be of help to the instrumentalist. Its core is the Nomic Postulate, according to which, given a domain, each conceptual frame has a unique strongest true hypothesis, i.e., the truth. Intermediate considerations apply both to the constructive empiricist and to the referential realist.

We can go even further. Given the proof of the Success Theorem, the following theorem is now easy to prove:

Forward Theorem

If CSH, which speaks of *remaining* more successful, is true, this implies the TA-hypothesis, that is, if *Y* is not closer to the nomic truth than *X*, (further) testing of CSH will sooner or later lead to an extra counterexample of *Y* or to an extra success of *X*.

In other words, 'so far proven to be more successful' can only be explained by the TA-hypothesis (Success Theorem) or by assuming that the comparative success hypothesis has not yet been sufficiently tested (Forward Theorem).

It is important to stress once more that the present section was based on the assumption of an observational vocabulary. I withdraw this assumption in the next two sections.

8. Intuitions of Scientists and Philosophers

The main point of Ch. 8.1 of ICR is that it is possible to give a 'dual foundation' of nomic truthlikeness and the corresponding methodology, which can be seen as an explication of some basic intuitions and practices of scientists. However, it turned out later that this

approach can also account for intuitions about the role of nonempirical, e.g. aesthetic, features of theories. Here I present the analysis with the option of explicating this additional intuition of scientists in mind. For this purpose I also leave room for a distinction between observational and theoretical terms.

The definitions of equal and more truthlikeness of Section 7 can be reformulated in terms of desirable and undesirable features of a theory. The starting point consists of properties of possibilities. A feature of a theory will be understood as a 'distributed' feature, that is, a property of all the possibilities that the theory admits. This leaves room for empirical features of theories, such as all its possibilities satisfying certain observational laws, but also for nonempirical features. For example, a theory is frequently called symmetric because all its possibilities show a definite symmetry. According to this definition, a feature of a theory can be represented as a set of possibilities, namely as the set of all possibilities that have the relevant property. This set then contains the set of all possibilities that the theory admits. Note that this means that we could say that a feature of a theory excludes (exactly) all possibilities that do not have that property.

It is obvious how we can formulate explicit definitions of desired, undesired, and remaining features in terms of the (logical) exclusion of desired and undesired possibilities: desired features are features that include all desired possibilities or, equivalently, that exclude only undesired possibilities; undesired features are features that include all undesired possibilities or, equivalently, that exclude only desired possibilities. All remaining features, as far as they can be represented as a subset of *CP*, exclude desired and undesired possibilities; that is, they do not include either all desired possibilities or all undesired ones. These are features about which we can be neutral, for which reason I call them neutral features. However, they will play no role in the following analysis.⁶ The three types of features are depicted in Figure 4.

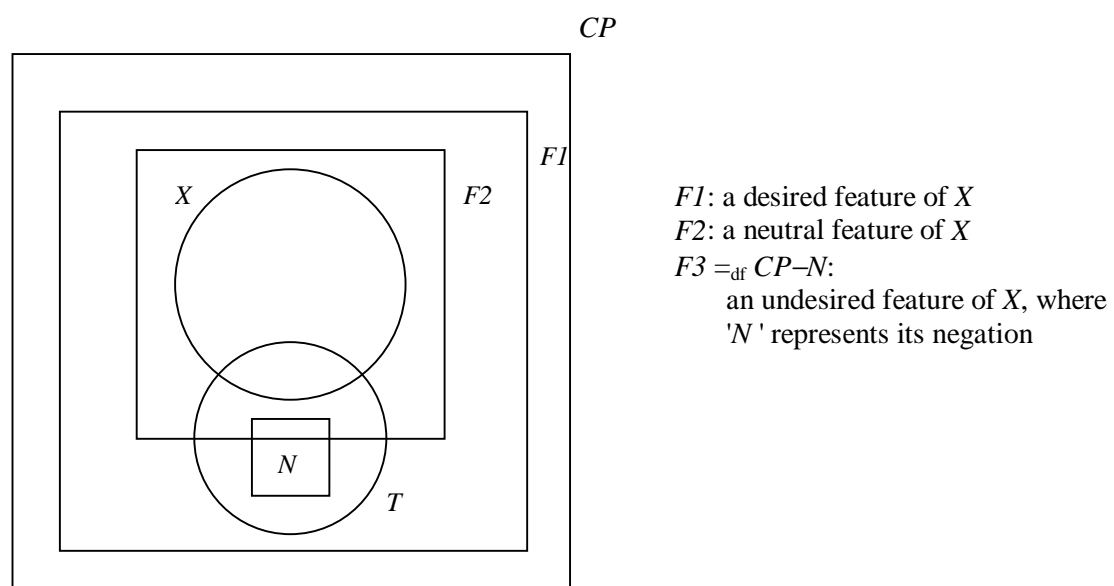


Figure 4: Three types of features

⁶ Popper has given a definition of 'closer to the truth' in terms of more true and fewer false consequences, that was acknowledged later (also by Popper himself) to be unsound. In terms of features Popper's mistake can be rephrased as an exceedingly broad understanding of undesired features: not only the features that are defined as undesired above, but also the neutral features fall under Popper's definition. For further analysis, see ICR, Section 8.1 and also (Zwart 1998/2001, Chapter 2), who has creatively reused part of Popper's intuitions (Chapter 6).

Note that a desired feature F of X is a true feature of X in the sense that not only X but also T is a subset of F , that is, the weak claim that may be associated with F , " $T \subseteq F$ ", is true. However, not only all undesired features of X are false in this sense but also all neutral features. The undesired features are false in a strong sense: they not only exclude some desired possibilities, but only such possibilities.

The following theorems can now easily be proved ⁷:

Equivalence theses

Y is at least as close to T as X iff

(UF) all undesired features of Y are also features of X (equivalent to (DP))

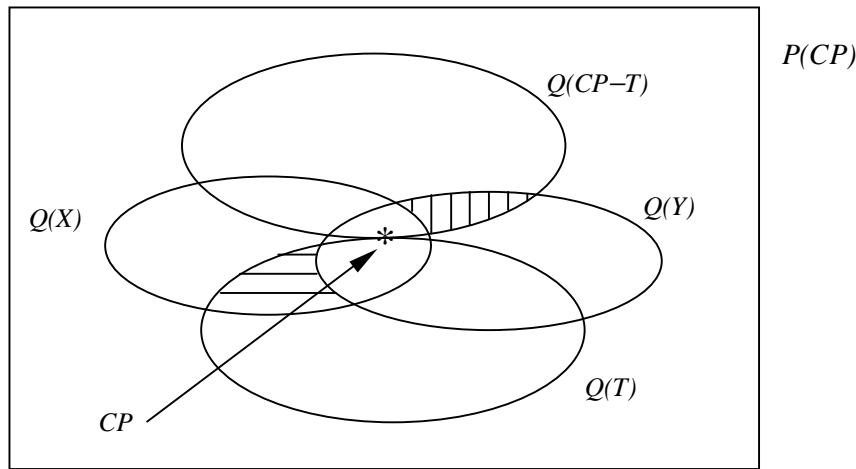
(DF) all desired features of X are also features of Y (equivalent to (UP))

Y is two-sidedly closer to T than is X iff

(UF) & (UF+) X has extra undesired features (equivalent to (DP+))

(DF) & (DF+) Y has extra desired features (equivalent to (UP+))

In Figure 5, 'at least as close to the truth' is depicted in terms of features. The rectangle now represents the 'universe' of all possibly relevant, distributed, features, and hence the powerset $P(CP)$ of CP (see Note 6). $Q(X)$ and $Q(Y)$ represent the set of features of X and Y , $Q(T)$ represents the set of desired features (features of T) and $Q(CP-T)$ represents the set of undesired features (the features of $CP-T$). Note that, $Q(T)$ and $Q(CP-T)$ have exactly one element in common, namely the tautology, which can be represented by CP .



⁷ Note for readers interested in the technical details. For proving these theorems it is advisable to introduce the set-theoretical interpretation of the universe of features and the set-theoretical characterization of the new clauses in terms of 'powersets' and 'co-powersets'. The powerset $P(X)$ of X is defined as the set of all subsets of X . The rectangle representing the 'universe' of all possibly relevant, distributed, features can now be interpreted as the 'powerset' $P(CP)$ of CP . Like a kind of mirror notion to that of powerset, the co-powerset $Q(X)$ of X is the set of all subsets of CP that include X , also called the supersets of X (within CP). $Q(X)$ then represents the features of X , $Q(T)$ the desired features and $Q(CP-T)$ the undesired features. Note that $Q(T)$ and $Q(CP-T)$ have exactly one set as common element, namely CP , that corresponds with the tautology, and that is of course included in the set of features of every theory. This results in the following formal translations of the four feature clauses:

(UF) $Q(Y) \cap Q(CP-T) \subseteq Q(X) \cap Q(CP-T)$ (UF+) $(Q(X) \cap Q(CP-T)) - Q(Y) \neq \emptyset$

(DF) $Q(X) \cap Q(T) \subseteq Q(Y) \cap Q(T)$ (DF+) $(Q(Y) \cap Q(T)) - Q(X) \neq \emptyset$

Proving the equivalence theses in terms of sets now becomes a nice exercise in 'set calculation'.

Figure 5: Y is at least as close to T as X , in terms of features.

(UF): $|||$ -area empty, (DF): \equiv -area empty

Notice the strong analogy between the logical form of (DP) (see the beginning of Section 7) and (DF) and between that of (DP+) and (DF+). The same goes for the logical form of (UP) and (UF), and of (UP+) and (UF+). The equivalences stated in the theorem, though, correspond in the reverse way: (DP) and (UF) are equivalent, as are (DP+) and (UF+), (UP) and (DF), (UP+) and (DF+). None of this is at all surprising, for undesired features could be defined in terms of desired possibilities and vice versa. Therefore it is in principle possible to reproduce the proof of the theses informally, clause by corresponding clause.⁸

On the basis of the equivalences, it follows that the two principal definitions can also be given in a mixed or 'dual' form, in terms of desired possibilities and desired features: (DP) and (DF) for at least as close to the truth, with the addition of (DP+) and (DF+) for (two-sidedly) closer to the truth. Roughly speaking, 'more truthlike' amounts to 'more desired possibilities and more desired features'. In my opinion, this resonates very much the intuitions of many scientists. We can strengthen this by taking into account that a desired possibility is a 'correct model' and a desired feature a 'true (general) consequence', where 'general' refers to all nomic possibilities. In this terminology, the dual conceptual foundation for nomic truthlikeness is most appealing to scientific common sense: 'more truthlike' amounts to 'more true (general) consequences and more correct models'. Moreover, it is now easy to see that 'more successful' amounts to 'more *established* true (general) consequences, i.e., general successes, and fewer *established* incorrect models, i.e., counterexamples', in line with the asymmetric definition, which is in fact also of a dual nature. Finally, dual nomic truthlikeness and the corresponding dual methodology leave room for nonempirical considerations in theory evaluation, such as aesthetic ones.

9. Epistemological Stratification of Nomic Truth Approximation

So far I have discussed objective features of theories in general. Of course there are different kinds of features and corresponding criteria. An obvious classification is the division into empirical and nonempirical features and criteria.

There are two main categories of empirical criteria of a theory, in accordance with the dual design above. I have already mentioned the question whether or not the theory implies a certain established observational law that, if so, can be explained or predicted by the theory. The entailment of an observational law can thus be conceived as an established desired observational feature of the theory. Observational laws are of course established by 'object induction' on properties recurring in repeated experiments. Instead of speaking of entailment or explanation and/or prediction of the theory, in what follows I will simply speak of explanation of such laws. Besides the 'explanation criterion' there is the 'instantial criterion', viz., the admission of an observed possibility, that is, the result of a particular experiment being an example or counterexample of the theory. So an observed possibility can be regarded as an established, desired observational possibility.

Assuming that empirical criteria are primary, relative to their possible aesthetic value, they are the only relevant criteria as long as only observational and no theoretical terms are involved. In other words, nonempirical features are only important if a (relative) distinction

⁸ Another note for readers interested in the technical details. Let me give, by way of example, a proof of the claim that (UF) entails (DP). Assume (UF) and let, contrary to (DP), x be a desired possibility admitted by X , that is, x belongs to $X \cap T$, and let x not be admitted by Y , hence belong to $T - Y$. Now $CP - \{x\}$ is a superset of Y , hence it represents a feature of Y which only excludes desired possibilities, viz. x , (and no undesired ones). Hence it is an undesired feature of Y , which should according to (UF) also be a feature of X , which rules out that x is a member of X . Q.e.d. All proofs are of this elementary nature.

between observational and theoretical terms can be made. I suppose that, in the present context, this distinction holds. Of course such a distinction between theoretical and observational terms leads to the distinction between an observational level of conceptual possibilities CP_o and a theoretical (cum observational) level of conceptual possibilities $CP = CP_t$. This distinction allows a precise definition of empirical versus nonempirical features to be formulated: features of the first kind exclude possibilities on the observational level, features of the second kind do not. Formally, e.g. for the second kind, a subset F of CP represents a nonempirical feature iff for all x in CP_o there is at least one y in CP_t such that y has x as its 'projection' in CP_o . This definition may suggest that nonempirical features of theories, in particular aesthetic ones, cannot be indicative of the empirical merits and prospects of a theory. However, by way of meta-induction, that is, inductive extrapolation or even generalization of a feature of certain theories to other ones, such features can come to be conceived as indicative in this respect. In this sense, aesthetic criteria may be seen as indirect empirical criteria, though formally quite different from the two categories of empirical criteria introduced above. From now on I shall speak only of empirical criteria (and features) in the direct sense explained above.

Truth approximation by means of empirical criteria can now be defined and founded on the basis of the following, easy to prove,

Combined Projection & Success theorem

If Y is closer to T than X then Y is at least as successful as X , almost in the sense of Sections 6 and 7, more precisely:

(DF-Success:) Explanatory clause

All established observational laws explained by X are also explained by Y (or: all established desired observational features of X are also features of Y)

(DP-Success:) Instantial clause

All observed examples of X are also examples of Y "unless X is lucky" (in other words: all observed counterexamples of Y are also counterexamples of X , "unless X is lucky").

The subclause "unless X is lucky" will be clarified shortly. The underlying assumption for the proof of this theorem is the correctness of the empirical data, that is to say, the observed possibilities and the observational laws that are (through an inductive leap) based on them, are correct.⁹

This theorem permits the functionality argumentation given in Section 7 to be generalized. Assume that theory Y at time t is (two-sidedly) more successful than X in the sense suggested above: not only are the two clauses fulfilled, but also Y explains at least one extra observational law and X has at least one extra observed counterexample (in other words: Y has an extra observed example). This evokes the comparative success hypothesis that Y will be lastingly more successful than X . This hypothesis is a neat empirical hypothesis of a comparative nature that can be tested by deriving and testing new test implications. As soon as this hypothesis has been sufficiently tested, in the eyes of some scientists, the rule of success can be applied, which means that they draw the conclusion that Y will remain more successful than X . It can be proved (recall the Forward Theorem of Section 7) that this is equivalent to concluding that the observational theory that follows from Y is closer to the observational truth To (the strongest true theory that can be formulated with the observational vocabulary, thus as a subset of CP_o) than X . But this conclusion is in its turn a good argument for the truth approximation (TA-)hypothesis on a theoretical level: Y is closer to the (theoretical) truth $T = T_t$ than X . In other words, the rule of success is functional for truth approximation. For this, three specific reasons have been given in Section 7 (for further

⁹ For the set-theoretical formulation of this theorem I refer to ICR Sections 7.3.3 and 9.1.1).

details, see ICR, p. 162, p. 214). They only need some qualification in view of the possibility of lucky hits, to which we now turn.

Whereas the explanatory clause is straightforward, the instantial clause is not, due to the possibility of lucky hits. It is interesting to study the latter in some detail. Let an I(nstantial)-similarity be an observed possibility that is admitted by both or neither theory, and let an I-difference be an observed possibility that is admitted by Y but not by X . Because of the 'one-many' character of the relation between the observational and theoretical levels of conceptual possibilities, a theory can have an observational feature on the observational level only if it has one on the theoretical level. The admission of an observational possibility on the theoretical level, though, cannot be based only on the admission of a suitable desired theoretical possibility, but can also be based on a suitable undesired theoretical possibility. If the observed example can be based on some admitted desired theoretical possibility, it may be called a real success of the theory. However, if the observed example can only be based on admitted undesired possibilities, it is some sort of *lucky hit* of that theory. For I-similarities there are all kinds of possibilities for this to occur, but it is not worth the effort of spelling them all out. I-differences, on the other hand, are very interesting. An I-difference can be based on a lucky hit of Y , in which case the DP+-clause, on the theoretical-cum-observational level, will not be verified and so the TA-hypothesis will not be confirmed (and therefore the reversed DP-clause is not falsified). Of course, if it is a real success of Y , the DP+-clause is verified, the TA-hypothesis is confirmed, and the reversed DP-clause is falsified. These possibilities are depicted in Figure 6.

In Figure 6, the observed example a is a real success of Y if there is a theoretical version in area 4 and it is a lucky hit if there is not, in which case there must be versions in 1 and 2. If (DF) [\Leftrightarrow (UP)] holds, area 2 is empty, so a must be a real success. It is clear that whether an extra success of Y is real or only apparent cannot be ascertained on the basis of the

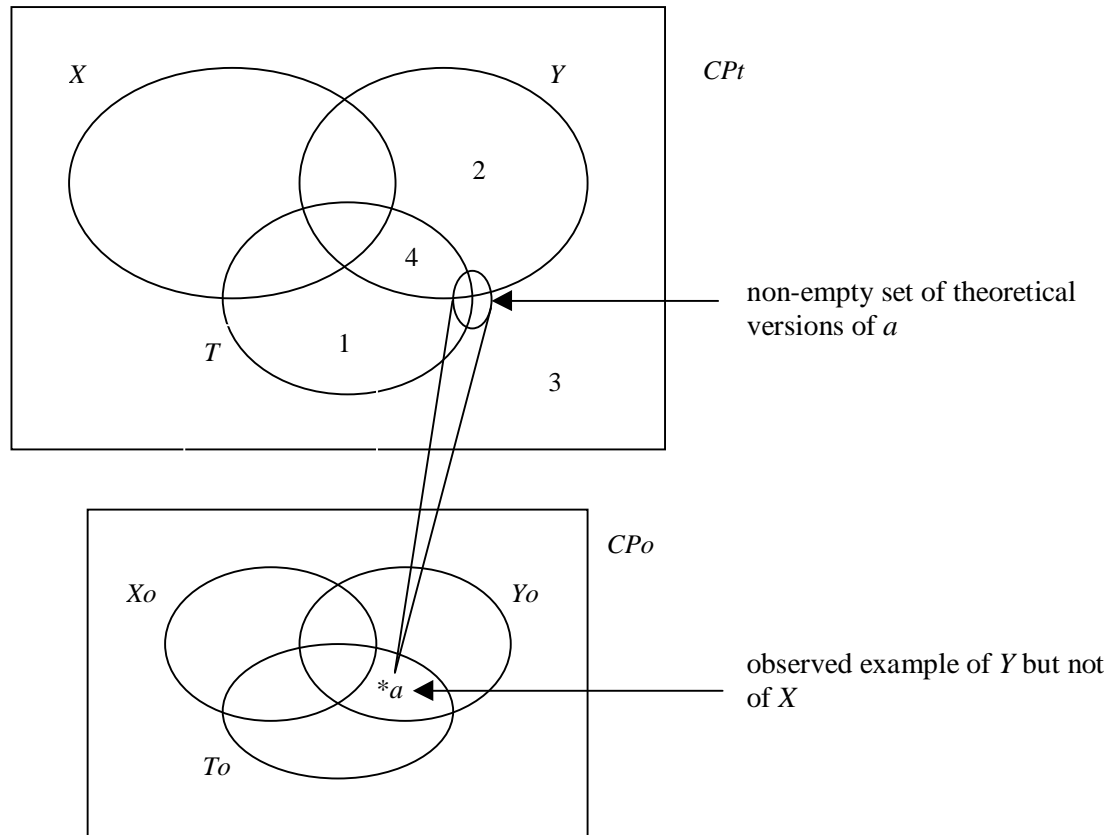


Figure 6: I-difference: observed example of Y (but not of X) as a real success or as a lucky hit. Further explanation in the text.

observed example. We can say, though, that if TAH (especially (DF)) is true, the example must be real. As said, in that case the DF+-clause is verified and TAH confirmed. Although this is not a completely circular confirmation, it is a '(DF)-laden' and therefore 'TAH-laden' confirmation. So an I-difference is not reliable as a (modest) 'signpost to the truth', even when it is correctly determined. This nature of I-differences makes it possible for realists who want to defend TAH in a concrete case, to relativize reversed I-differences: after all, an instantial success of *X* that is a counterexample of *Y* could be a lucky hit on the part of *X*. This is the condition mentioned in the instantial clause.

As suggested before, not only empirical criteria may play a role in theory choice, but also nonempirical criteria, that is, criteria in terms of logical, conceptual or aesthetic features. In a recent paper (Kuipers, 2002) a *formal-cum-naturalistic* analysis is given of the relation between beauty, empirical success, and truth. It supports the findings of James McAllister in his inspiring *Beauty and revolution in science* (1996), by explaining and justifying them. First, scientists are essentially right regarding the usefulness of aesthetic criteria for truth approximation, provided they conceive of them as less hard than empirical criteria. Second, the aesthetic criteria of the time, the 'aesthetic canon', may well be based on 'aesthetic (meta-) induction' regarding (distributed) nonempirical features of paradigms of successful theories which scientists have come to appreciate as beautiful. Third, they can play a crucial, dividing role in scientific revolutions. Since aesthetic criteria may well be wrong, they may retard empirical progress and hence truth approximation in the hands of aesthetic dogmatists but not in the hands of aesthetically flexible, 'revolutionary' scientists.

The truth approximation analysis also affords an opportunity to reconsider the nature of descriptive and explanatory research programs. Such programs presuppose, by definition (see SiS, Ch. 1) a domain, a problem, and a core idea, including a vocabulary, to solve that problem. A *descriptive research program* uses an observational conceptual frame, and may either exclusively aim at one or more true descriptions (as for example in most historiography), or it may also aim at the true (observational) theory in the following specific way. In this nomological type of descriptive program the goal of a true theory is supposed to be achieved exclusively by establishing observational laws. Given that this requires (observational) inductive jumps, it is plausible to call such a descriptive program an *inductive research program*. It is easy to see that such programs 'approach the truth by induction'. The micro-step of the HD method may be applied for the establishment of observational laws, resulting in true descriptions which either falsify the relevant general observational hypothesis or are partially derivable from it. According to the basic definition of 'more truthlike', assuming that accepted observational laws are true, any newly accepted observational law guarantees a step in the direction of the true theory. For it is easy to verify that if $S(t)$ and $S(t')$ indicate the strongest accepted law at time t and t' later than t , respectively, $S(t')$ is closer to T than $S(t)$. Hence, inductive research programs are relatively safe strategies of truth approximation: as far as the inductive jumps happen to lead to true accepted laws, the approach not only makes truth approximation plausible, it even guarantees it.

Let me now turn to the explication of the nature of explanatory or theoretical programs, which are by definition of a nomological nature. An *explanatory program* may or may not use a theoretical vocabulary. Even (nomic) empiricists can agree that it is directed at establishing the true observational theory. If there are theoretical terms involved, the referential realists will add that it is also directed at establishing the referential truth. The theory realist will add to this that it is even directed at establishing the theoretical truth. Scientists working within such a program will do so by proposing theories respecting the hard core as long as possible, but hopefully not at any price. They will HD evaluate these theories separately and comparatively. RS directs theory choice and is trivially functional for empirical progress. Moreover, although that rule is demonstrably functional for all distinguished kinds of nomic truth approximation, it cannot guarantee a step in the direction of the relevant truth, even assuming correct data. Though the basic notions of successfulness and truthlikeness are sufficient to give the above characterization of the typical features of explanatory research programs, they usually presuppose refined means of comparison, which are presented in Part IV of ICR.

IV REFINED TRUTH APPROXIMATION

To keep this synopsis within reasonable limits, I have chosen not to give a detailed impression of the last part of ICR, Chh. 10-12. Brief indications of the chapters will have to suffice. Ch. 10 introduces another sophistication of the basic approach: it accounts for the fact that progress is frequently made by new theories that introduce new mistakes, something which is excluded according to basic truth approximation. In Ch. 11 this refinement allows some real-life illustrations of (potential) truth approximation, one from physics and another from economics. Moreover, in Ch. 12 it is shown that there are also quantitative versions of refined truth approximation, based upon distances between structures.

10. Refinement of Nomic Truth Approximation

The study of truthlikeness and truth approximation is completed by introducing a second major sophistication accounting for a fundamental feature of most theory improvement, viz., new theories introduce new mistakes, but mistakes that are in some way less problematic than the mistakes they replace. This refinement is introduced in a qualitative way by taking into account that one incorrect model may be more similar, or 'more structurelike' to a target model than another. This leads to refined versions of nomic truthlikeness and truth approximation, with adapted conceptual foundations. It is argued, and illustrated by the Law of Van der Waals, that the frequently and variously applied method of 'idealization and successive concretization', propagated by Nowak, is a special kind of (potential) refined nomic truth approximation (in this respect, see also Kuipers, forthcoming a). Combining the present sophistication with that of Section 9, one obtains explications of stratified refined nomic truthlikeness and truth approximation.

11. Examples of Potential Truth Approximation

Two sophisticated examples illustrate that the final analysis pertains to real-life, theory-oriented, empirical science. The first example shows that the successive theories of the atom, called 'the old quantum theory', viz., the theories of Rutherford, Bohr, and Sommerfeld, are such that Bohr's theory is closer to Sommerfeld's than Rutherford's. Here, Bohr's theory is a (quantum) specialization of Rutherford's theory, whereas Sommerfeld's is a (relativistic) concretization of Bohr's theory. This guarantees that the nomic truth, if not caught by the theory of Sommerfeld itself, could have been a concretization of the latter. In both cases, Sommerfeld would have come closer to the truth than Bohr and Rutherford. The second example illustrates a non-empirical use of the idealization and concretization methodology, viz., aiming at (approaching) a provable interesting truth. In particular, it is shown that the theory of the capital structure of firms of Modigliani and Miller is closer to a provable interesting truth than the original theory of Kraus and Litzenberger, of which the former is a 'double' concretization.

12. Quantitative Truthlikeness and Truth Approximation

Here the prospects for quantitative versions of actual and nomic truthlikeness are investigated. In the nomic refined case there are essentially two different ways of corresponding quantitative truth approximation, a non-probabilistic one, in the line of the qualitative evaluation methodology, and a probabilistic one, in which the truthlikeness of theories is estimated on the basis of a suitable probability function. As stressed in Ch. 3 of ICR, probabilistic methodological reasoning, notably about confirmation, is already rather

artificial, although it is used by scientists to some extent. However, quantitatively measuring the distance between theories and between their successes is in most cases even more artificial and, moreover, rare. Hence, the quantitative accounts of (non-) probabilistic truth approximation, notably that of Niiniluoto (1987), are presented with many reservations.

13. Conclusion: Constructive Realism

Recall that Section 1 introduces the main epistemological positions, instrumentalism, constructive empiricism, referential realism, constructive (theory) realism and essentialist (theory) realism. In the course of ICR the following conclusion could provisionally be drawn at the end of Part III and could be further strengthened in Part IV:

The instrumentalist methodology provides good reasons for the transition of epistemological positions from instrumentalism to constructive realism. Here, the intermediate step from constructive empiricism to referential realism turned out to be the hardest one, whereas the step from constructive to essentialist realism had to be rejected.

The rest of ICR Ch. 13 presents the main lines of the resulting favorite epistemological position of *constructive realism*. It is a conceptually relative, hence non-essentialist, nomic truth approximation version of theory realism, accounting for objective truths. They can be approached by an intersubjective method, viz., the evaluation methodology for theories, in which the role of (truth-)testing of hypotheses primarily concerns testing test implications of theories as well as testing comparative success and truth approximation hypotheses of theories.

The term 'constructive realism' has earlier been used by Giere (1985), and my conception of it is rather similar, except that I include in it, of course, truth approximation, whereas Giere still focuses on the true/false dichotomy, but he fully recognizes the nomic aim of theorizing. With respect to truth approximation, my position is rather similar to that of Niiniluoto (1987, see in particular Section 4.3). The main difference between my and his position, besides my primarily qualitative versus his straightforward quantitative approach, is my emphasis on the nomic aim of theorizing. In sum, constructive realism reflects the combination of their deviating strengths by emphasizing *nomic truth approximation* as opposed to the *actual truth-value* of theories.

Ch. 13 deals more in particular with the acceptance (as true) of three types of hypotheses or claims, that is, three types of induction, viz. observational, referential and theoretical induction, and with the formation of observation terms. The resulting metaphysical nature of scientific research is depicted, together with portraits of real and fictitious scientists. It concludes with a discussion of metaphors for empirical science research, and concludes that the map metaphor, rather than the mirror or the net metaphor, is to be preferred, although it is certainly not perfect.

As already mentioned in the introduction, there arises in ICR a clear picture of scientific development, with a short-term and a long-term dynamic. In the former there is a severely restricted role for confirmation and falsification; the dominant role is played by (the aim of) empirical progress, and there are serious prospects for observational, referential and theoretical truth approximation. Hence, regarding this short-term dynamic, the scientist's intuition that the debate among philosophers about instrumentalism and realism has almost no practical consequences can be explained and justified. The long-term dynamic is enabled by (observational, referential and theoretical) inductive jumps, after 'sufficient confirmation', providing the means to enlarge the observational vocabulary in order to investigate new domains of reality. In this respect, a consistent instrumentalist epistemological attitude seems difficult to defend, whereas constructive realism seems the most plausible.

I would like to conclude this synopsis by explaining the reason for this last claim, originating from the crucial role played by observational and theoretical induction in the

construction and determination of terms and, hence, the long-term dynamic in science. Besides the formation of observation terms by straightforward explicit definition, observational induction may provide the necessary and sufficient empirical conditions, e.g., in the form of existence and uniqueness requirements, for explicitly defining new observation terms. The quantitative notions of pressure and temperature are examples. Such new terms are unambiguously, hence intersubjectively, applicable, and they (may be supposed to) refer if they enable new observational inductions. Besides implying referential induction, theoretical induction may provide the necessary and sufficient conditions, e.g., in the form of existence and uniqueness requirements, for applying theoretical terms, that is, for identifying theoretical entities and for measuring theoretical attributes. For example, the detection of electrons, and the measurement of their mass and charge, are based on such inductions. In this way, theoretical terms (may be supposed to) become referring and unambiguously, hence intersubjectively, applicable. In other words, theoretical terms can be essentially transformed into new observation terms by appropriate theoretical and/or referential induction. Of course, together with earlier or elsewhere accepted observation terms, they can be used for new cases of observational induction. Moreover, with other ones, they will play a crucial role in the separate and comparative evaluation of new theories introducing new theoretical terms, dealing with (partially) new domains, starting another round of the 'empirical cycle'. For an epistemological instrumentalist it is difficult to account for this long-term dynamic in a consistent way. However, for a constructive realist, focusing on nomic truth approximation, this is easy, ironically enough, in particular when he is prepared to replace the falsificationist methodology by the instrumentalist methodology.

University of Groningen

Department of Theoretical Philosophy

Aweg 30

9718 CW Groningen

The Netherlands

T.A.F.Kuipers@philos.rug.nl

<http://www.philos.rug.nl/personae/kuipers.html>

Appendix 1: Table of Contents ICR

From Instrumentalism to Constructive Realism: On some relations between Confirmation, Empirical Progress, and Truth Approximation

	Contents, Foreword
Chapter 1	General Introduction: Epistemological Positions
1.1.	Four perspectives on theories
1.2.	The four main epistemological questions
1.3.	The main epistemological and methodological claims
1.4.	Preliminaries and a survey of cognitive structures
Part I	Confirmation
	Introduction to Part I
Chapter 2	Confirmation by the HD Method
2.1.	A qualitative theory of deductive confirmation
2.2.	Ravens, emeralds, and other problems and solutions
2.3.	Acceptance of hypotheses
Chapter 3	Quantitative Confirmation, and its Qualitative Consequences
3.1.	Quantitative confirmation
3.2.	Qualitative consequences
3.3.	Acceptance criteria
	Appendix 1: Corroboration as inclusive and impure confirmation
	Appendix 2: Comparison with standard analysis of the raven paradox
Chapter 4	Inductive Confirmation and Inductive Logic
4.1.	Inductive confirmation
4.2.	The continuum of inductive systems
4.3.	Optimum inductive systems
4.4.	Inductive analogy by similarity and proximity
4.5.	Universal generalizations
Part II	Empirical Progress
	Introduction to Part II
Chapter 5	Separate Evaluation of Theories by the HD Method
5.1.	HD evaluation of a theory
5.2.	Falsifying general hypotheses, statistical test implications, and complicating factors
Chapter 6	Empirical Progress and Pseudoscience
6.1.	Comparative HD evaluation of theories
6.2.	Evaluation and falsification in the light of truth approximation
6.3.	Scientific and pseudoscientific dogmatism
Part III	Basic Truth Approximation
	Introduction to Part III
Chapter 7	Truthlikeness and Truth Approximation
7.1.	Actual truthlikeness
7.2.	Nomic truthlikeness
7.3.	Actual and nomic truth approximation
7.4.	Survey of bifurcations
7.5.	Novel facts, crucial experiments, inference to the best explanation, and descriptive research programs
Chapter 8	Intuitions of Scientists and Philosophers
8.1.	Conceptual foundations of nomic truth approximation
8.2.	Truthlikeness and the correspondence theory of truth

8.3.	Explicating dialectical concepts
Chapter 9	Epistemological Stratification of Nomic Truth Approximation
9.1.	Theoretical and substantial nomic truth approximation
9.2.	Referential truth approximation
9.3.	Rules of inference, speculations, extensions, and explanatory research programs
9.4.	Epistemological positions reconsidered
Part IV	Refined Truth Approximation
	Introduction to Part IV
Chapter 10	Refinement of Nomic Truth Approximation
10.1.	Structurelikeness
10.2.	Refined nomic truthlikeness and truth approximation
10.3.	Foundations of refined nomic truth approximation
10.4.	Application: idealization & concretization
10.5.	Stratified refined nomic truth approximation
Chapter 11	Examples of Potential Truth Approximation
11.1.	The old quantum theory
11.2.	Capital structure theory
Chapter 12	Quantitative Truthlikeness and Truth Approximation
12.1.	Quantitative actual truthlikeness and truth approximation
12.2.	Quantitative nomic truthlikeness
12.3.	Quantitative nomic truth approximation
Chapter 13	Conclusion: Constructive Realism
13.1.	Main conclusions
13.2.	Three types of induction
13.3.	Formation of observation terms
13.4.	Direct applicability of terms
13.5.	The metaphysical nature of scientific research
13.6.	Portraits of real and fictitious scientists
13.7.	Reference and ontology
13.8.	Truth definitions and truth criteria
13.9.	Metaphors
	Notes, References, Index of Names, Index of Subjects

Appendix 2: Outline Table of Contents SiS

Structures in Science: Heuristic Patterns based on Cognitive Structures. An advanced textbook in neo-classical philosophy of science

Contents, Foreword

Part I Units of Scientific Knowledge and Knowledge Acquisition

- 1 Research programs and research strategies
- 2 Observational laws and proper theories

Part II Patterns of Explanation and Description

- 3 Explanation and reduction of laws
- 4 Explanation and description by specification

Part III Structures in Interlevel and Interfield Research

- 5 Reduction and correlation of concepts
- 6 Levels, styles, and mind-body research

Part IV Confirmation and Empirical Progress

- 7 Testing and further separate evaluation of theories
- 8 Empirical progress and pseudoscience

Part V Truth, Product, and Concept Approximation

- 9 Progress in nomological, design, and explicative research
- 10 Design research programs

Part VI Capita Selecta

- 11 Computational philosophy of science
- 12 The structuralist approach to theories
- 13 'Default-norms' in research ethics

Suggestions for further reading, Exercises, Notes, References, Index of Names, Index of Subjects

Appendix 3: Acronyms

A-difference/-similarity	Aesthetic difference/similarity
CD-hypothesis	Correct Data hypothesis
CP	set of Conceptual Possibilities
CSH	Comparative Success Hypothesis
d-confirmation	deductive confirmation
DF/UF	Desired Features / Undesired Features (clause)
DF-success	Desired Features success (clause)
DP/UP	Desired Properties / Undesired Properties (clause)
DP-success	Desired Properties success (clause)
E-difference/-similarity	Explanatory difference/similarity
GTC	General Testable Conditional
GTI	General Test Implication
HD (evaluation, method, testing)	Hypothetico-Deductive (evaluation, method, testing)
IC	Initial Condition(s)
ICR	<i>From Instrumentalism to Constructive Realism</i>
I-difference/-similarity	Instantial difference/similarity
ITI	Individual Test Implication
LMC	Logico-Mathematical Claim
<i>Mod(H)</i>	the set of models of H
MP	Modus (Ponendo) Ponens
MT	Modus (Tollendo) Tollens
PC	Principle of Content
PCE	Principle of Comparative HD evaluation
PD	Principle of Dialectics
PI	Principle of Improvement
PIRP	Principle of Improvement by Research Programs
PSE	Principle of Separate HD evaluation
PT	Principle of (Falsifiability or) Testability
<i>p</i> -(non-)zero	(non-)zero probability
RE	Rule of Elimination
RS	Rule of Success
SiS	<i>Structures in Science</i>
STR	Special Theory of Relativity
TA-hypothesis / TAH	Truth Approximation Hypothesis
UI	Universal Instantiation

References

- Compton, K. (1988). 0-1 Laws in Logic and Combinatorics. In: I. Rival (ed.), *Proceedings 1987 NATO Adv. Study Inst. on Algorithms and Order*, pp. 353-383. Dordrecht: Reidel.
- Earman, J. (1992). *Bayes or Bust. A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT-press.
- Festa, R. (1993). *Optimum Inductive Methods. A Study in Inductive Probability Theory, Bayesian Statistics and Verisimilitude*. Dordrecht: Kluwer Academic Publishers.
- Festa, R. (1999) Bayesian Confirmation. In: M.C. Galavotti and A. Pagnini (eds.), *Experience, Reality, and Scientific Explanation*, pp. 55-87. Dordrecht: Kluwer Academic Publishers.
- Fitelson, B. (1999). The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science*, Supplement to Volume **66** (3), S362-S378.
- Giere, R. (1985). Constructive Realism. In: P. Churchland and C. Clifford (eds.), *Images of Science*, pp. 75-98. Chicago: The University of Chicago Press.
- Groot, A. de, (1961/1969). *Methodologie*, Den Haag: Mouton, 1961; translated as: *Methodology*. New York: Mouton, 1969.
- Grove, A. Halpern, J. and Koller, D. (1996). Asymptotic Conditional Probabilities: The Unary Case. *The Journal of Symbolic Logic*, **61** (1), 250-275.
- Hempel, C. (1966) *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall.
- Hodges, W. (1986). Truth in a Structure. *Proceedings of the Aristotelian Society, New Series*, Vol. **86**, London.
- Howson, C. & Urbach, P. (1989). *Scientific Reasoning: the Bayesian Approach*. La Salle: Open Court.
- Kemeny, J. (1953). A Logical Measure Function. *The Journal of Symbolic Logic*, (**18**) 4, 289-308.
- Kuipers, T. (1978). *Studies in Inductive Probability and Rational Expectation*. *Synthese Library*, vol. 123. Dordrecht: Reidel.
- Kuipers, T. (1998). Pragmatic Aspects of Truth Approximation. In: P. Weingartner, G. Schurz and G. Dorn (eds.), *The Role of Pragmatics in Contemporary Philosophy*. (Proceedings of the 20th International Wittgenstein-Symposium, August 1997), pp. 288-300. Vienna: Hölder-Pichler-Tempsky.
- Kuipers, T. (1999). Abduction Aiming at Empirical Progress or Even at Truth Approximation, Leading to a Challenge for Computational Modelling. In: J. Meheus and T. Nickles (eds.), *Scientific Discovery and Creativity*, special issue of *Foundations of Science* **4** (3), 307-323.
- Kuipers, T. (2000/ICR). *From Instrumentalism to Constructive Realism. On Some Relations between Confirmation, Empirical Progress, and Truth Approximation*. *Synthese Library*, vol. 287. Dordrecht: Kluwer Academic Publishers.
- Kuipers, T. (2001/SiS). *Structures in Science. Heuristic Patterns Based on Cognitive Structures. An Advanced Textbook in Neo-Classical Philosophy of Science*. *Synthese Library*, vol. 301. Dordrecht: Kluwer Academic Publishers.
- Kuipers, T. (2001). The Logic of Progress in Nomological, Explicative and Design Research. Ch. 9 of Kuipers (2001/SiS), pp. 255-264.
- Kuipers, T. (2002). Beauty, a Road to The Truth. *Synthese*, **131** (3), 291-328.
- Kuipers, T. (forthcoming a). Empirical and Conceptual Idealization and Concretization. The Case of Truth Approximation. Forthcoming in (English and Polish editions of) *Liber Amicorum for Leszek Nowak*.
- Kuipers, T. (forthcoming b). Inference to the Best Theory. Kinds of Induction and Abduction. Forthcoming in the Proceedings of the ESF-workshop *Induction and Deduction in the Sciences*, Vienna, July, 2002.
- Lakatos, I. (1970/1978). Falsification and the Methodology of Scientific Research Programmes. In: I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, pp. 91-196. Cambridge: Cambridge University Press; reprinted in Lakatos (1978), pp. 8-101.

- Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*, eds. J. Worrall and G. Currie. Cambridge: Cambridge University Press.
- Laudan, L. (1977). *Progress and its Problems*. Berkeley: University of California Press.
- Maher, P. (this volume). Qualitative Confirmation and the Ravens Paradox.
- McAllister, J. (1996). *Beauty and Revolution in Science*. Ithaca: Cornell University Press.
- Milne, P. (1996). $\text{Log}[P(h/eb)/P(h/b)]$ is the one true measure of confirmation, *Philosophy of science*, **63**, 21-26.
- Mura, A. (1990). When Probabilistic Support is Inductive, *Philosophy of Science*, **57**, 278-289.
- Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.
- Panofsky, W. and Phillips, M. (1955, 1962²). *Classical Electricity and Magnetism*. London: Addison-Wesley.
- Popper, K. (1934/1959). *Logik der Forschung*. Vienna, 1934; translated as: *The Logic of Scientific Discovery*. London: Hutchinson, 1959.
- Popper, K. and Miller, D. (1983). A Proof of the Impossibility of Inductive Probability. *Nature*, **302**, 687-688.
- Salmon, W. (1969). Partial Entailment as a Basis for Inductive Logic. In: N. Rescher (ed.), *Essays in honor of Carl G. Hempel*, pp. 47-82. Dordrecht: Reidel.
- Schlesinger, G. (1995). Measuring Degrees of Confirmation. *Analysis*, **55** (3), 208-212.
- Zwart, S. (1998/2001). *Approach to The Truth. Verisimilitude and Truthlikeness*. Dissertation Groningen. Amsterdam: ILLC-Dissertation-Series-1998-02. Revised version: *Refined Verisimilitude*, Synthese Library, vol. 307. Dordrecht: Kluwer Academic Publishers.